Invariant Causal Routing for Governing Social Norms in Online Market Economies

Xiangning Yu
Tianjin University
Tianjin, China

Qirui Mi
Institute of Automation, Chinese Academy of Sciences
Beijing, China

Xiao Xue*
Tianjin University
Tianjin, China

Haoxuan Li
Peking University
Beijing, China

Yiwei Shi
University of Bristol
Bristol, United Kingdom

Xiaowei Liu
Tianjin University
Tianjin, China

Mengyue Yang*
University of Bristol
Bristol, United Kingdom

Abstract

Social norms are stable behavioral patterns that emerge endogenously within economic systems through repeated interactions among agents. In online market economies, such norms—like fair exposure, sustained participation, and balanced reinvestment—are critical for long-term stability. We aim to understand the causal mechanisms driving these emergent norms and to design principled interventions that can *steer* them toward desired outcomes. This is challenging because norms arise from countless micro-level interactions that aggregate into macro-level regularities, making causal attribution and policy transferability difficult. To address this, we propose **Invariant Causal Routing (ICR)**, a causal governance framework that identifies policy–norm relations stable across heterogeneous environments. ICR integrates counterfactual reasoning with invariant causal discovery to separate genuine causal effects from spurious correlations and to construct *interpretable, auditable policy rules* that remain effective under distribution shift. In heterogeneous agent simulations calibrated with real data, ICR yields more stable norms, smaller generalization gaps, and more concise rules than correlation or coverage baselines, demonstrating that causal invariance offers a principled and interpretable foundation for governance.

CCS Concepts

• **Social and professional topics** → *Governmental regulations*; • **Applied computing** → **Online shopping**; • **Computing methodologies** → **Simulation evaluation**; *Multi-agent planning*.

*Corresponding authors.

Conference'17, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Keywords

Multi-Agent Systems, Economic Markets, Social Norms, Causal Inference, PNS, Interpretability

ACM Reference Format:
Xiangning Yu, Qirui Mi, Xiao Xue, Haoxuan Li, Yiwei Shi, Xiaowei Liu, and Mengyue Yang. 2026. Invariant Causal Routing for Governing Social Norms in Online Market Economies. In . ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Social norms are commonly defined as stable patterns of behavior that emerge within groups through repeated interactions. They delineate acceptable individual actions while sustaining social order and cooperation at the collective level [8, 16]. Their binding force rests on mutual expectations: individuals comply because they anticipate others will do the same [29]. A defining property of norms is therefore their *stability*—they persist over long time horizons, withstand environmental shocks, and display consistent adherence across groups and contexts. We focus on understanding how social norms arise in economic systems and how governance actors can intentionally *steer* the formation of social norms toward desirable collective outcomes.

However, the formation of social norms in large-scale adaptive systems is far from transparent. Countless local interactions aggregate into macro-level regularities [29], making it difficult to attribute outcomes to specific actions or interventions. This opacity becomes critical when the governance actor (the online market) seeks to steer norm formation through intervention. Governance rarely dictates behavior directly; instead, it shapes *structural conditions*—such as subsidies, fee rates, and exposure or pricing rules—to elicit self-organized responses that crystallize into collective norms. However, interventions often yield divergent results, effective in one context yet ineffective in another, because outcomes depend less on implementation quality than on behavioral cues and confounders including visibility, sanction, reciprocity, and imitation.

These difficulties highlight several research challenges. First, norms emerge from decentralized interactions, making causal responsibility opaque. Second, interventions face multistability and

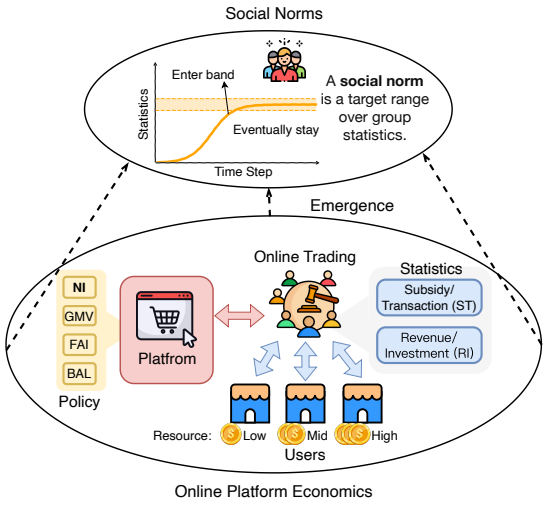

Figure 1: Social norms in onlineeconomies.

path dependence, as identical policies may lead to divergent equilibria under different initial conditions. Third, heterogeneity and asymmetry in populations amplify disparities: the same cue may elicit different reactions across groups, and uneven visibility or sanctioning further skews outcomes. Finally, even when structural relations remain intact, distributional shifts in confounders or exogenous variables distort correlations, weaken extrapolation, and raise the cost of trial-and-error.

Existing research has primarily emphasized natural emergence or aggregate outcomes in multi-agent coordination, focusing on what works "on average." However, such approaches do not address the governance-relevant question: *why does a given intervention succeed in one context but fail in another?* If distributional drift is the root cause, the remedy lies not in increasingly complex correlational modeling but in uncovering **causal structures invariant across environments**.

This paper advances the claim that governance should prioritize identifying **invariant causal factors** that determine success or failure under distributional shift, rather than comparing surface-level strategies. To this end, we adopt **causal invariance** as a guiding principle and employ the **Probability of Necessity and Sufficiency (PNS)** to provide counterfactual evidence. PNS formalizes success as occurring *if and only if* a given intervention is applied, thereby enabling the identification of **invariant causal routes** that remain effective across initial conditions, perturbations, and random seeds. Compared with correlation-based heuristics, PNS yields stronger out-of-distribution robustness and interpretability.

We define *norm achievement* as a threshold event where key macro statistics (e.g., revenue-to-investment and subsidy-to-transaction ratios) enter and persist within a target range, capturing long-term stability while tolerating short-term fluctuations. To operationalize this idea, we introduce a three stage causal framework that (i) identifies policies effective only when applied, revealing those invariant across environments; (ii) compiles them into a concise rule table of the form "if context ψ, then apply intervention θ," emphasizing robustness and minimal redundancy; and (iii) traces the causal pathway from governance actions to fundamental factors and eventual norm achievement, providing interpretable explanations.

We validate this framework in an online marketplace with heterogeneous users, where subsidies and exposure policies vary while user strategies remain fixed. Norm achievement is assessed by whether macro statistics enter and persist within the target range, allowing direct evaluation of governance effectiveness.

This paper advances a causal perspective on governance by moving beyond average-effect optimization toward discovering invariant causal relations that determine policy success under distributional shifts. Our main contributions are:

(1) **Algorithmic Innovation – Invariant Causal Routing (ICR).** We propose a novel three-stage framework that integrates causal inference and rule-based policy learning. ICR identifies *causal routes*—policy-to-norm relations that remain valid across heterogeneous environments—and formulates them into a minimal, auditable rule list.
(2) **Interpretability and Causal Accountability.** Grounded in the Probability of Necessity and Sufficiency (PNS), ICR provides transparent causal guarantees, distinguishing genuine policy effects from confounded correlations and offering human-readable explanations of causal mechanisms.
(3) **Empirical Validation under Distribution Shift.** Through heterogeneous-agent simulations calibrated with real-world data, ICR maintains stable norm attainment under out-of-distribution conditions, achieving smaller generalization gaps and higher robustness than correlation-based and heuristic baselines.

Together, these contributions show that **causal invariance** offers a principled basis for designing stable, interpretable, and transferable governance strategies in complex socio-economic systems.

2 Related Work

The related literature spans three complementary strands.

Social Norms. Research on social norms examines their spontaneous emergence without central control, explained through expectation–compliance frameworks [8], stochastic evolutionary games [50, 51], and multi-agent social laws [43, 44]. Social psychology and experimental economics study their definition, maintenance, and measurement [7, 20, 27, 33], while recent LLM-based work shows that linguistic interaction can itself yield conventions and biases [12, 19]. Yet prior studies lack (i) operational criteria for detecting norm formation in long-term data and (ii) causal tools linking interventions to macro norms [8, 11, 12, 20, 27, 33, 41, 42, 52]. We define norm attainment as system-level indicators stabilizing within a tolerance band, enabling causal analysis along the chain intervention → micro responses → macro norms.

From Correlational to Causal Policy Governance. Conventional policy evaluation relies on correlations with limited causal interpretability. A/B testing and observational designs estimate average treatment effects [3, 13], while offline estimators such as propensity scores, IPW, and DR [14, 37], and macro designs like difference-in-differences or synthetic control [1], struggle with heterogeneity [4, 23, 28, 34, 48] and distribution shifts [6, 22, 35]. They often

yield broad rather than context-specific insights [5, 13], while offline evaluation faces bias–variance and robustness challenges [24, 46]. Deep RL, bandit, and learning-to-rank methods improve short-term outcomes [15, 25, 40, 45] but remain opaque and hard to audit [39]. Bridging interpretability, stability, and causal validity thus remains an open challenge.

OOD Prediction and PNS. PNS formalizes the causal responsibility of an intervention in potential-outcome terms. Given treatment $A \in \{0, 1\}$ and potential outcomes $Y(A)$, PNS $= \Pr\{Y(1) = 1,\ Y(0) = 0\}$ measures the probability that an event occurs only under treatment [34, 47]. Causal approaches to out-of-distribution (OOD) generalization assume domain-invariant mechanisms. Some explicitly constrain causal graphs [17, 21, 32, 36, 38, 53], while others treat invariance as a representation-learning objective, such as IRM and its extensions in game-theoretic, variance-penalized, and nonlinear settings [2, 18, 27, 30]. Unlike these constraints, *PNS-based invariant learning* [49] integrates PNS directly into training, minimizing necessity–sufficiency mismatch via paired counterfactuals to yield representations that are both causally decisive and stable across domains.

3 Preliminaries and Problem Formulation

3.1 Social Norms in Online Market

Social norms are stable, self-enforcing behavioral regularities that emerge endogenously through repeated interactions among heterogeneous agents. In online market economies, such norms manifest as persistent collective patterns—**fair exposure**, **sustained participation**, or **balanced reinvestment**—rather than externally imposed rules. They are characterized not by precise equilibria, but by long-run *bands* of system-level indicators (e.g., exposure share, activity level, reinvestment ratio) that remain within tolerance ranges over time.

In this work, we treat social norm formation as a measurable, long-term event: a system achieves a norm when its trajectory statistics enter and *persist* within a specified tolerance band. This operationalization allows for direct causal analysis and accommodates short-term fluctuations.

Formally, let $\phi : \mathcal{X} \to \mathbb{R}^d$ denote bounded continuous statistics (e.g., group-wise Revenue/Investment ratio or Subsidy/Transaction ratio). After a burn-in period T_{burn}, the window-T average is:

$$\bar{\phi}_T = \frac{1}{T} \sum_{t=T_{\text{burn}}+1}^{T_{\text{burn}}+T} \phi(X_t), \tag{1}$$

where $\phi(X_t)$ captures d-dimensional group metrics at time t.

Definition 3.1 (Social Norm Band). Given a target vector $\eta \in \mathbb{R}^d$ and tolerance $\varepsilon \in (0, \infty)^d$, the *social norm band* is

$$\mathcal{S}_\varepsilon = \prod_{j=1}^{d} [\eta_j - \varepsilon_j,\ \eta_j + \varepsilon_j]. \tag{2}$$

A run under policy θ *attains the norm* if there exists $T_0 < \infty$ such that $\bar{\phi}_T \in \mathcal{S}_\varepsilon$ for all $T \geq T_0$.

Definition 3.2 (Social norm attainment). Under platform strategy θ, a run *attains a social norm* if

$$\exists\, T_0 < \infty \quad \forall\, T \geq T_0 : \quad \bar{\phi}_T \in \mathcal{S}_\varepsilon. \tag{3}$$

In other words, the time-averaged statistics eventually stay within the band, and exact convergence to a single point is unnecessary.

Distance convention. We use the sup norm $\|v\|_\infty = \max_j |v_j|$ on $\mathbb{R}^d$ and the induced distance to a set:

$$\text{dist}_\infty(x, S) \ = \ \inf_{y \in S} \|x - y\|_\infty. \tag{4}$$

For the rectangular band $\mathcal{S}_\varepsilon = \prod_{j=1}^{d} [\eta_j - \varepsilon_j, \eta_j + \varepsilon_j]$, this reduces to

$$\text{dist}_\infty(x, \mathcal{S}_\varepsilon) = \max_{1 \leq j \leq d} \Big\{ \max \big\{ (\eta_j - \varepsilon_j) - x_j,\ x_j - (\eta_j + \varepsilon_j),\ 0 \big\} \Big\}. \tag{5}$$

Existence of feasible limits. Under mild and standard assumptions (compactness and Feller continuity), Lemma B.1 (App. B) ensures the set of invariant measures $\mathcal{I}(P^{(\theta)})$ is nonempty and compact. Consequently, the *asymptotic signature set*

$$\Sigma(P^{(\theta)}, \phi) = \left\{ \int \phi \, d\pi : \pi \in \mathcal{I}(P^{(\theta)}) \right\} \subset \mathbb{R}^d \tag{6}$$

is nonempty and compact, and every subsequential limit of $(\bar{\phi}_T)$ lies in $\Sigma(P^{(\theta)}, \phi)$ by Theorem C.4 (App. C). Thus, Σ organizes the feasible long-run profiles of $(\bar{\phi}_T)$.

Proposition 3.3 (Norm Feasibility Criterion). *If $\Sigma(P^{(\theta)}, \phi) \cap \mathcal{S}_\varepsilon \neq \varnothing$, the social norm is* feasible*, meaning there exists a stationary regime whose averages satisfy the band. Conversely, if $\Sigma(P^{(\theta)}, \phi) \cap \mathcal{S}_\varepsilon = \varnothing$, then there exists a constant $\delta > 0$ such that*

$$\liminf_{T \to \infty} \text{dist}_\infty \left(\bar{\phi}_T, \mathcal{S}_\varepsilon \right) \geq \delta \quad \textit{almost surely.} \tag{7}$$

That is, stable attainment is impossible if all invariant averages lie outside the band.

The existence result makes the event well-posed; feasibility depends on whether the band intersects $\Sigma(P^{(\theta)}, \phi)$. We do not require $\bar{\phi}_T$ to converge to a single limit—"eventual stay" suffices. Further discussion on convergence and ergodicity appears in App. D.

3.2 System Setup and Dynamics

We consider a platform ecosystem comprising a platform and multiple heterogeneous user groups that repeatedly interact over time. The platform follows a fixed objective $\theta \in \Theta$ and influences user behavior through policy levers such as subsidies, fee rates, and exposure thresholds. The joint state of the system evolves as a homogeneous Markov chain:

$$\{X_t\}_{t \geq 0}, \qquad X_{t+1} \sim P^{(\theta)}(\cdot \,|X_t), \tag{8}$$

on a compact state space $\mathcal{X}$. Each θ corresponds to a distinct governance regime—e.g., GMV growth (GMV), fairness (FAI), balanced growth–fairness (BAL), or user welfare/retention (UW)—with a normal-intervention (NI) baseline as reference.

3.3 Problem Definition

The governance objective is to identify causal mappings from platform policies θ to long-run social-norm outcomes and determine which mappings remain *invariant* across heterogeneous environments or distributional shifts. This reframes governance from *average-effect optimization* to *causal invariance*—discovering factors and routes that *consistently cause* norm attainment regardless of confounders. We therefore investigate three key research questions:

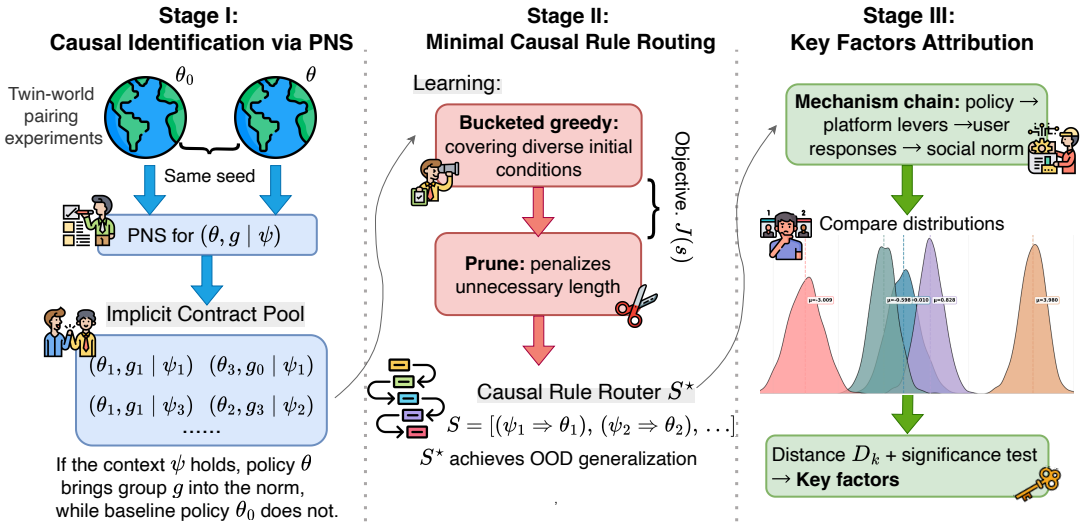

Figure 2: Three-stage framework for discovering causal strategies in social norm formation. Stage I tests causal effects by identifying implicit contracts $(\theta, g \mid \psi)$ where policy θ enables group g to reach the norm under context ψ. Stage II learns a compact rule router $S^\star$ selecting effective strategies for robust norm attainment. Stage III explains success by attributing norms to key factors where user responses differ from the baseline θ_0.

(1) **RQ1 (What: causal impact).** How do different platform objectives *causally influence* the emergence and stabilization of social norms across user groups?
(2) **RQ2 (How: invariant causal routing).** Under changing or out-of-distribution (OOD) conditions, what is the *shortest and most stable routing policy*—a compact mapping from context to action—that maintains consistent causal effects on norm attainment?
(3) **RQ3 (Why: causes of norm divergence).** What underlying *mechanisms and structural factors* explain why different groups or contexts evolve toward distinct social norms, and which are necessary or sufficient for these divergences?

4 Method: Three-Stage Causal Governance

We propose a three-stage causal governance framework addressing RQ1–RQ3: **Stage (I) Causal identification** estimates probabilistic necessity–sufficiency (PNS) to test whether a strategy θ causes a group g to attain its norm band under context ψ. **Stage (II) Invariant causal routing** learns a minimal first-match rule list mapping contexts to actions, maximizing norm attainment under distribution shift while ensuring causal stability. **Stage (III) Key factors attribution** contrasts θ with its baseline to reveal structural and behavioral factors that drive norm stabilization or divergence.

4.1 Stage I: Causal Identification via PNS

We define when a platform strategy θ is *causally responsible* for a group g meeting its target band, relative to a baseline θ_0, through the **Probability of Necessity and Sufficiency (PNS)**.

Definition 4.1 (Potential Outcome and PNS). Let $Y_g(\theta) = B^{(g)}(\theta) \in \{0, 1\}$ denote the potential outcome for group g under intervention θ, where $B^{(g)}(\theta) = 1$ indicates the group meets the target band. Following Pearl's counterfactual framework [34], for a context predicate ψ on the initial state, the Probability of Necessity and Sufficiency (PNS) is the probability that replacing the baseline strategy θ_0 with θ enables group g to enter the norm band.

$$\mathrm{PNS}_g(\theta \mid \psi) = \Pr\left(Y_g(\theta) = 1, Y_g(\theta_0) = 0 \mid \psi = 1\right). \tag{9}$$

This probability captures both the *necessity* (group does not meet the target under θ_0) and *sufficiency* (group meets the target under θ) of the platform strategy.

PNS Identification. We estimate PNS using paired runs on the same seed. No additional exogeneity or monotonicity assumptions are needed due to the paired design. We test whether a strategy θ is *causally responsible* for enabling group g to attain its social-norm band under a given context ψ.

Estimation. Because both potential outcomes are observed under paired runs, PNS is point-identified by the unbiased estimator

$$\widehat{\mathrm{PNS}}_g(\theta \mid \psi) = \frac{1}{N_\psi} \sum_{e:\psi(X_e)=1} \mathbf{1}\left\{Y_{g,e}(\theta) = 1,\ Y_{g,e}(\theta_0) = 0\right\}, \tag{10}$$

where N_ψ is the number of seeds with $\psi(X_e) = 1$. Confidence intervals see App. E.

Definition 4.2 (Implicit Contract). Given thresholds $n_{\min} \in \mathbb{N}$ and $\tau_{\mathrm{pns}} \in [0, 1]$, we call $(\theta, g \mid \psi)$ an *implicit contract* if it satisfies both:

$$\text{(support)} \quad N_\psi \geq n_{\min}, \tag{11}$$

$$\text{(accountability)} \quad \widehat{\mathrm{PNS}}_g(\theta \mid \psi) \geq \tau_{\mathrm{pns}}, \tag{12}$$

i.e., the clause has sufficient support across initial conditions and its PNS confidence exceeds the target level.

Interpretation. An implicit contract is a reliable "if–then" promise in the social-norm sense: *if the context ψ holds for an initial condition, then applying θ is causally sufficient to bring group g into norm compliance, whereas baseline behavior would fail.*

Outcome. Stage I thus yields a sound pool of *implicit contracts* $(\theta, g \mid \psi)$ that satisfy both support and accountability, forming the candidate set for Stage II.

4.2 Stage II: Minimal Causal Rule Routing

From Stage I we obtain a set of *implicit contracts* $(\theta, g \mid \psi)$ with certified PNS. We assemble them into a compact, ordered decision list:

$$S = [(\psi_1 \Rightarrow \theta_1),\ (\psi_2 \Rightarrow \theta_2),\ \ldots,\ (\psi_{|S|} \Rightarrow \theta_{|S|})], \tag{13}$$

which applies the *first* matching rule to an episode with context X_0; otherwise we fall back to θ_0.

Covering diverse initial conditions (bucketed scoring). To explicitly expand coverage over heterogeneous initial conditions, we partition the initial contexts into buckets $\mathcal{B} = \{b\}$ (e.g., by quantiles or clustering on X_0) with empirical weights w_b. For each candidate rule i we compute its *bucketed marginal coverage* $m_{i,b}$ (the fraction of bucket-b episodes newly covered under first-match order) and its certified causal gain $\mathrm{PNS}_{g,i,b}$ per group g within bucket b. This favors rules that both *reach new initial-condition mass* and *deliver certified gains* where they apply.

Objective. We maximize a bucketed objective that rewards union coverage of initial conditions and penalizes unnecessary length:

$$J(S) = \sum_{b \in \mathcal{B}} w_b \sum_{i=1}^{|S|} m_{i,b} \cdot \Big(\sum_{g} w_g \,\mathrm{PNS}_{g,i,b} \Big) \;-\; \lambda\,|S|. \tag{14}$$

Here $m_{i,b}$ accounts for first-match semantics (only *new* coverage counts), ensuring the list grows coverage across buckets rather than overfitting a few.

Learning: bucketed greedy + prune. We learn a short first-match rule list on validation environments as follows: (i) rank candidates by $\sum_g w_g \,\mathrm{PNS}_{g,i}$; (ii) iteratively add the rule with the largest *bucketed marginal* improvement in $J(S)$—only not-yet-covered contexts count—and enforce a *coverage safeguard* $\min_b \mathrm{Cov}_b(S) \geq \tau_{\mathrm{cov}}$; (iii) run a backward prune that removes any rule whose deletion reduces J by at most τ_{prune}. The resulting $S^\star$ covers diverse initial-condition buckets with few rules; we then evaluate $S^\star$ on held-out `test` seeds. See Algorithm 1.

Outcome. Stage II yields a parsimonious router $S^\star$ with invariant causal effects. With PNS-filtered clauses and stable rule gains across buckets, $S^\star$ generalizes reliably under distribution shift.

4.3 Stage III: Key Factors Attribution

Having identified effective contracts and assembled a rule router, we now seek to explain *why* certain strategies succeed while others fail under the *same initial conditions*. Stage III isolates the mechanism factors that differ when the PNS event—success under θ but failure under θ_0—occurs.

Matched comparison. For an implicit contract $(\theta, g \mid \psi)$, we focus on the same set of episodes satisfying $\psi(X_e) = 1$, representing comparable starting conditions for group g. Within this matched set, we contrast the system's internal responses under the target and baseline policies. Formally, for each lever factor f_k,

$$S_k^{(\theta)} = \{\, f_k(e,\theta) : \psi(X_e) = 1 \,\}, \qquad S_k^{(\theta_0)} = \{\, f_k(e,\theta_0) : \psi(X_e) = 1 \,\}. \tag{15}$$

This captures how the same initial state evolves differently under alternative interventions.

Distributional divergence. We quantify the induced change on each factor using a normalized distance metric

$$D_k = \mathrm{Dist}\big(S_k^{(\theta)}, S_k^{(\theta_0)}\big), \tag{16}$$

where Dist can be Wasserstein-1. The metric, normalization, and estimation details are provided in App. G.2.

Identifying key factors. A factor f_k is deemed *key* if (i) $D_k \geq \theta_{\mathrm{dist}}$, and (ii) the difference is statistically significant under the test and correction procedures described in App. G.2. These factors represent the levers through which θ causally alters user or platform behavior relative to θ_0.

Outcome. Stage III yields ranked key factors that separate successful from failed interventions under identical conditions, offering interpretable causal levers for understanding and transferring social-norm formation.

Algorithm 1 LearnRouter (Greedy + Buckets + Prune)

Require: Candidates $C = \{(\psi, \theta)\}$; splits train/val/test; buckets B with weights w_b; penalty λ; coverage threshold τ_{cov}; prune tol. τ_{prune}; max rules K
Ensure: Ordered router S

$S \leftarrow [\,]$; for $b \in B$: $\mathrm{Cov}[b] \leftarrow \emptyset$ *(first-match covered set)*
while $|S| < K$ **do**
  for each $c \in C \setminus S$ **do**
    $\mathrm{ncov}[c] \leftarrow \sum_b w_b \cdot \mathrm{NewCov}(c, b, \mathrm{Cov}[b]; \mathrm{train})$
    $\mathrm{gain}[c] \leftarrow \sum_b w_b \cdot \mathrm{Gain}(c, b; \mathrm{val})$
    $\mathrm{score}[c] \leftarrow \mathrm{ncov}[c] \cdot \mathrm{gain}[c] - \lambda$
  end for
  $c^\star \leftarrow \arg\max_c \mathrm{score}[c]$;
  if $\mathrm{score}[c^\star] \leq 0$ **then**
    break
  end if
  append $c^\star$ to S; update all $\mathrm{Cov}[b]$
  if $\min_b \mathrm{BucketCov}(S, b; \mathrm{val}) < \tau_{\mathrm{cov}}$ **then**
    add best remaining rule that raises uncovered-bucket coverage with positive score
  end if
end while
for each rule r in S from last to first **do**
  if $\mathrm{Val}(S \setminus \{r\}; \mathrm{val}) \geq \mathrm{Val}(S; \mathrm{val}) - \tau_{\mathrm{prune}}$ **then**
    remove r
  end if
end for

return S

5 Experiments

We evaluate the proposed three-stage causal governance framework in a simulated online marketplace *ecosystem* calibrated with real-world data from the 2022 *Survey of Consumer Finances* (SCF) [10]. To address RQ 1, we estimate PNS via twin-world pairing to quantify

the causal impact of platform interventions on norm attainment. For RQ 2, we learn a minimal first-match router $S^\star$ that generalizes across seeds and initial economic regimes, verifying causal invariance under distribution shift. For RQ 3, we perform mechanism attribution by contrasting endogenous platform levers and user responses across causal routes to reveal why certain interventions induce or fail to sustain stable social norms.

5.1 Experimental Setup

Platform Intervention Goals. We consider five operational objectives for the online market: (**NI**) a fixed normal-intervention baseline; (**GMV**) maximizing transactions or gross merchandise volume growth; (**FAI**) improving exposure fairness to ensure equal opportunity across user groups; (**BAL**) pursuing a balanced objective combining GMV growth and fairness; and (**UW**) maximizing long-run ecosystem welfare and user retention.

Agent Heterogeneity. The ecosystem comprises a *platform agent* and multiple *user agents*. The platform adjusts *subsidies*, *fee/take-rate tiers*, *exposure thresholds*, and off–transaction spending, while users choose their *investment share* and *activity level*. Users are grouped by resources (*low*, *mid*, *high*), forming four distinct behavioral strategies within the system.

Initial Conditions. To cover diverse regimes, we evaluate five initial configurations (IC1–IC5): baseline equilibrium, conservative, aggressive, balanced-liquidity, and robust-inequality economies (parameters in App. F).

Seed Shift as OOD. Unless otherwise noted, the structural equations and noise distribution are fixed across runs. Different random seeds are treated as a weak out-of-distribution (OOD) shift in the realization of exogenous noise and initial conditions, following $X_{t+1} = f_\theta(X_t, \xi_t)$ with $\xi_t \sim P(\xi)$. Training, validation, and test sets use disjoint seeds, while twin-world comparisons share the same noise realization within each seed e for (θ, θ_0).

Target Social Norms. We evaluate four social norms defined by two aggregate ratios: Subsidy/Transaction (ST) and Revenue/Investment (RI). For ST, **ST-1** represents low subsidy intensity where users rely mainly on organic traffic (laissez-faire), and **ST-2** denotes a regressive pattern—higher subsidies for low-resource users and lower for high-resource ones—stabilizing early growth but risking over-correction. For RI, **RI-1** indicates upward mobility, where low-resource users achieve higher ROI than incumbents, while **RI-2** captures concentration or entrenchment, where top users compound advantage and newcomers struggle to accumulate. Precise social norm band ranges are given in App. F.

Training Methodology. Agents are trained with *MADDPG* in a continuous, interdependent multi-agent environment with behaviors (investment/effort) and dynamic platform levers (subsidies/fees/exposure). Detailed hyperparameters are in App. G.

Evaluation Statistics. We measure norm emergence using two trajectory statistics $\phi_j(X_t)$: ϕ_1 (Subsidy-to-Transaction ratio, ST) capturing platform subsidy intensity, and ϕ_2 (Revenue-to-Investment ratio, RI) as a group-wise ROI proxy. After burn-in, we compute the window average $\bar{\phi}_{j,T}$ over the last T=50 steps (one step = one year).

Ablation Study Baselines. We train routers on `train` seeds and evaluate on disjoint `test` seeds across IC1–IC5; seed splits and baseline configurations are in App. G.1. Our baselines:

- **PNS+Greedy**: Stage I supplies PNS-certified clauses; rank candidates by mean target-conjunction PNS on `train`, add rules greedily, and stop when the validation objective $J(S)$ has no further strict improvement (i.e., $\Delta J \le \epsilon_{\text{imp}}$). The score includes a length penalty λ_{len}.
- **PNS+Greedy (pruned)**: same selection, followed by L_0-style pruning using the same $J(S)$; remove a rule if the decrease is at most τ_{prune}.
- **Corr+Greedy (Pearson)**: replace PNS with the Pearson correlation r (treatment vs. target-band attainment) for ranking; otherwise identical to the greedy procedure.
- **Corr+Greedy (Pearson, pruned)**: correlation-based greedy selection with the same L_0 pruning.
- **Coverage-Driven**: rank by unconditional target-band coverage on treated episodes (ignoring baseline failure); favors breadth over causal guarantees.
- **Coverage+Corr Hybrid**: rank by $\alpha\,\overline{\text{Cov}} + (1-\alpha)\,\bar{r}$, with the same greedy and pruning steps.
- **Majority Router**: for each (θ_0, norm), choose the single task that wins the most paired episodes on `train`; output one unconditional rule (targets=`ALL`).
- **Random Router**: uniformly random permutation of candidate tasks with a fixed RNG seed s (default s in App. G.2); a weak, non-causal baseline.

Metrics. $\mathbf{PNS_{Target}}$ (Train/Test): measures *causal* band entry for the target group(s) after switching policy—originally out of band, now in band. **Coverage** (Train/Test): plain hit rate under the current distribution—what fraction are in band *now* under θ; ignores where they would be under θ_0. No counterfactuals. **Rules** ↓: number of clauses in the learned first-match router (smaller is better). **Gen. Gap** ↓: train–test drop in causal effectiveness,

$$\text{Gap} \;=\; \text{PNS}^{\text{train}}_{\text{Target}} \;-\; \text{PNS}^{\text{test}}_{\text{Target}}. \tag{17}$$

Perf. ↑: overall test score that rewards causal effectiveness and basic attainability while penalizing overfitting and complexity,

$$\text{Perf} \;=\; w_{\text{pns}}\,\text{PNS}^{\text{test}}_{\text{Target}} + w_{\text{cov}}\,\text{Coverage}^{\text{test}} - w_{\text{gap}}\,\text{Gap} - \lambda_{\text{len}}{\cdot}\text{Rules}_{\text{norm}}. \tag{18}$$

where $\text{Rules}_{\text{norm}} = \text{Rules}/K$, and K is a normalization constant used only for the length penalty (set to the maximum router length per table; K=80 in our ablations).

5.2 Experimental Results

5.2.1 *Experiment 1: Existence and Rule Discovery of Invariant Causal Routing.*

Visual evidence of norm existence. Using one training-set seed, we record the final 50 steps after the system has fully evolved. Fig. 3 plots these trajectories across four norms (rows) and five tasks (columns). Group trajectories (low/mid/high-resource) relative to shaded bands confirm that multiple norms are stably attained under specific tasks: (i) **ST-1** is sustained under GMV; (ii) **ST-2** emerges under BAL; (iii) **RI-1** appears clearly under FAI; and (iv) **RI-2** requires

Figure 3: Last-50-year trajectories of ϕ_1 (ST) and ϕ_2 (RI) with norm bands (per objective × group). Shaded regions of the same color indicate the band within which individuals of the corresponding group are expected to remain under the associated social norm. When all trajectories lie within their respective same-colored shaded regions, the social norm is fully attained.

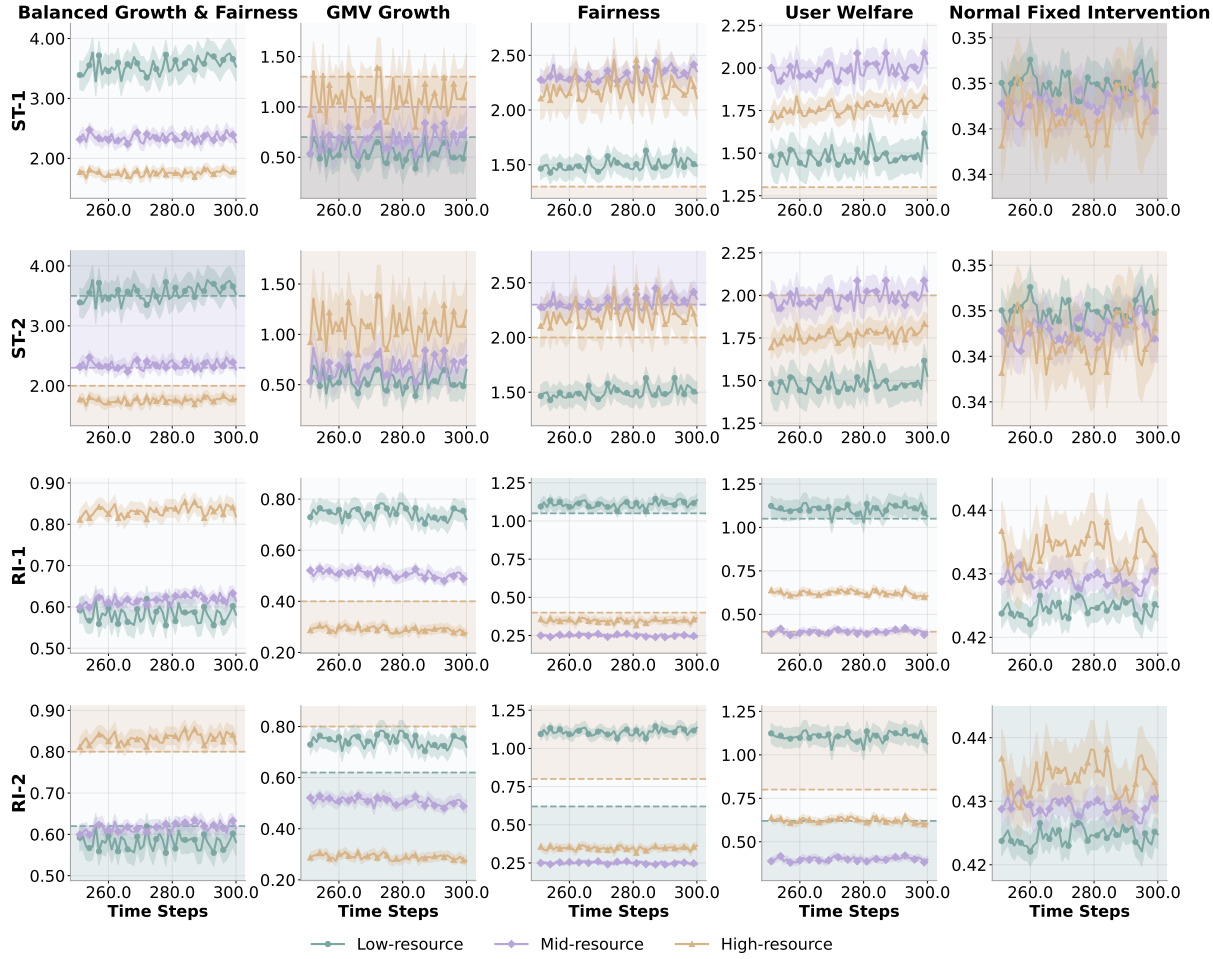

Table 1: Representative PNS-significant causal routes selected by Stage I and Stage II. Entries report PNS with 95% binomial confidence intervals as *value [lower, upper].*

Norm	Baseline	Current	Group	PNS
RI-1	NI	FAI	ALL	0.966 [0.904, 0.993]
RI-1	GMV	FAI	ALL	0.981 [0.899, 1.000]
RI-1	BAL	FAI	ALL	0.962 [0.893, 0.992]
ST-1	BAL	GMV	ALL	0.857 [0.722, 0.933]
ST-1	FAI	GMV	ALL	0.887 [0.774, 0.947]
ST-1	UW	GMV	ALL	0.905 [0.779, 0.962]
RI-2	NI	BAL	ALL	0.833 [0.735, 0.900]
RI-2	FAI	BAL	ALL	0.864 [0.761, 0.927]

BAL, while NI fails to hold bands. **These results provide direct evidence that social norms manifest within a simulated online market.**

Invariant causal routes. For example, *RI-1: NI→FAI (ALL)* with PNS = 0.966 means that replacing normal intervention (NI) by a fairness-oriented strategy (FAI) *causes* all groups to attain the RI-1 band with probability 0.966 under the covered contexts. Likewise, *ST-1: BAL→GMV (ALL)* shows that switching from balanced growth–fairness (BAL) to a GMV-focused objective *causes* the subsidy-to-transaction ratio (ϕ_1) to enter the ST-1 band.

PNS-certified clauses provide **causal validity** (RQ1): they quantify how strategies differentially induce norm attainment across groups. The Stage-II router provides **actionable parsimony** (RQ2): a short, feasible rule list that generalizes across seeds/contexts while maintaining high compliance.

Toward generalization. A key strength of this design is that PNS-based selection eliminates confounding and certifies invariance through twin-world estimation. As a result, the identified routes generalize across seeds without re-tuning—unlike correlation-based approaches that drift with confounder shifts. The next section

Table 2: Ablation results on disjoint train/test seeds (Perf recomputed with K=80). Higher ↑ is better; lower ↓ is better.

Method	$\text{PNS}_{\text{train}}$ ↑	$\text{Cov}_{\text{train}}$ ↑	PNS_{test} ↑	Cov_{test} ↑	Rules ↓	$\text{Rules}_{\text{norm}}$ ↓	Gap ↓	Perf. ↑
PNS+Greedy	**0.989**	**0.990**	**0.953**	**0.966**	24	0.300	0.036	0.862
PNS+Greedy (pruned)	0.972	0.981	0.931	0.938	**12**	**0.150**	0.041	**0.883**
Corr+Greedy (Pearson)	0.805	0.958	0.741	0.931	46	0.575	0.064	0.600
Corr+Greedy (Pearson, pruned)	0.742	0.840	0.677	0.796	18	0.225	0.065	0.627
Coverage-Driven	0.622	0.944	0.565	0.915	48	0.600	0.057	0.449
Coverage+Corr Hybrid	0.416	0.564	0.384	0.550	80	1.000	**0.032**	0.114
Majority Router	0.396	0.540	0.353	0.519	20	0.250	0.043	0.307
Random Router	0.294	0.393	0.251	0.354	60	0.750	0.043	0.042

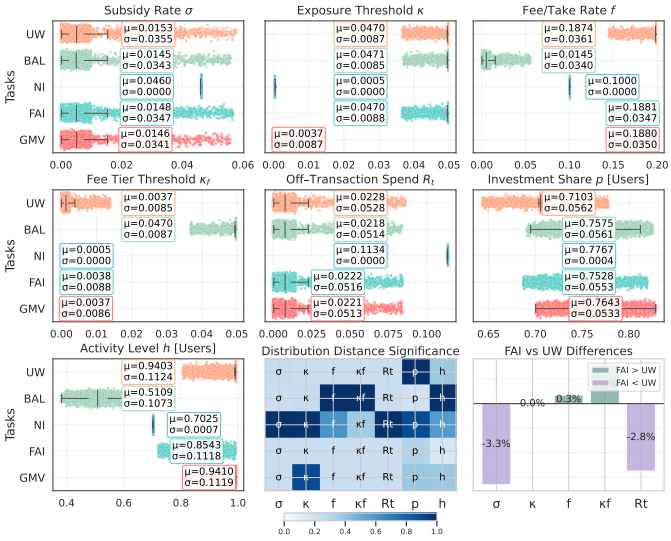

Figure 4: Key factors attribution for invariant causal routes.

(Exp. 5.2.2) evaluates this generalization explicitly, showing that PNS-based routers maintain strong compliance and small gaps on disjoint test seeds.

5.2.2 *Experiment 2: Ablations of Invariant Causal Routing.*

Results and analysis. **PNS+Greedy** and **PNS+Greedy (pruned)** dominate on test-time causal effectiveness and generalization under seed shift, reflecting the benefit of selecting clauses with certified necessity–sufficiency. Pruning improves parsimony with negligible loss—yielding shorter, cleaner rule lists while preserving causal validity. **Corr+Greedy (Pearson)** and **Corr+Greedy (Pearson, pruned)**—are weaker under shift, showing larger gaps and lower causal attainment, consistent with sensitivity to spurious associations. **Coverage-Driven** and **Coverage+Corr Hybrid** trade causality for breadth: they can raise unconditional hit rate but require long lists and lack counterfactual accountability, which limits robustness. **Majority Router** and **Random Router** serve as lower bounds on both effectiveness.

Overall, our PNS-guided router *realizes invariant causal routing* under distribution shift: it preserves certified causal effects across disjoint seeds while achieving strong test-time performance with compact rule lists.

5.2.3 *Experiment 3: Interpretability and Plausibility of Invariant Causal Routing.*

Platform Levers and User Responses. By stage III, we compare the *endogenous* distributions of (i) **platform levers**: subsidy rate (σ), exposure threshold (κ), commission (f), fee–tier threshold (κ_f)[1], and off–transaction spend (R_t)[2], and (ii) **user responses**: investment share (p) and activity level (h). Each task fixes a policy mapping π^{task}. The platform applies π^{task}, users react with (p, h), and the system converges to a *social norm*. Fig. 4 shows the per–task distributions for levers and responses, highlights levers with significant cross–task differences, and provides the *FAI* vs. *UW* mean deltas (bottom–right).

Why switching tasks produces these PNS routes. For example:
BAL → GMV ⇒ ST-1 (Route 4 in Table 1). Fig. 4 further shows that κ (exposure threshold) and κ_f (fee–tier threshold) decline while the commission f (platform take rate) increases. Lower thresholds make it easier to gain exposure and to qualify for lower fee tiers, which drives h (activity) upward. Volume rises but subsidies σ remain flat, so per-transaction subsidy intensity is low across groups, again producing *ST-1* .

NI → FAI ⇒ RI-1 (Route 1 in Table 1). Under *FAI*, the platform reduces off–transaction spend R_t and lowers the fee–tier threshold κ_f. The lower κ_f directly expands access to low-fee tiers, allowing more low-resource users to qualify for a lower *realized* take rate f and improving their relative inclusion (RI); meanwhile, high-resource users' relative advantage narrows. This pattern is consistent with *RI-1*.

PNS separates even subtle cases (FAI vs. UW). In the lever panels, FAI and UW almost overlap. The bottom–right plot shows mean gaps of only a few percent: FAI has slightly higher κ_f (fee–tier threshold) and f (commission), while UW has slightly higher R_t (off–transaction spend) and σ (subsidy rate). Even such small, systematic shifts alter behavior at the margin: under FAI, the activity level h is slightly lower but the investment share p is higher, whereas under UW the reverse holds. PNS distinguishes these cases by tracing the full pathway—*fixed policy → endogenous levers* $(\sigma, \kappa, f, \kappa_f, R_t)$ *→ user responses* (p, h) *→ social norms*—even when visual differences are minimal.

These results thus answer RQ3 by revealing interpretable, causally plausible routes of norm formation.

[1]Minimum volume to qualify for a lower take-rate tier; distinct from the take-rate f.
[2]$R_t \equiv R_t = \rho_t \cdot Y_t$, where ρ_t is the spend share and Y_t is aggregate GMV. Off–transaction spend covers marketing/ops, trust & safety, and tooling.

6 Conclusion

This work introduced **Invariant Causal Routing (ICR)**, a causal governance framework for discovering stable policy–norm relations under distribution shift. ICR integrates causal identification, compact rule learning, and key factor attribution to isolate genuine causal effects and construct interpretable governance strategies. Through heterogeneous agent simulations calibrated with real data, ICR demonstrates improved norm stability, smaller generalization gaps, and more concise rules compared with correlation or coverage based baselines. These results highlight the value of causal invariance as a foundation for reliable and interpretable governance in complex systems.

Beyond this study, the causal routing perspective opens several promising directions. Future work will extend ICR to multi-level governance with adaptive agents, explore causal meta-learning for dynamic environments, and validate the framework using empirical data from large-scale digital platforms. We hope this work contributes to building principled, transparent, and transferable approaches for governing social norms in evolving socio-economic ecosystems.

References

[1] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. 2010. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association* 105, 490 (2010), 493–505.

[2] Kartik Ahuja, Karthikeyan Shanmugam, Kush Varshney, and Amit Dhurandhar. 2020. Invariant risk minimization games. In *International Conference on Machine Learning*. PMLR, 145–155.

[3] Joshua D Angrist and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

[4] Susan Athey and Guido W Imbens. 2017. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic perspectives* 31, 2 (2017), 3–32.

[5] Susan Athey and Stefan Wager. 2021. Policy learning with observational data. *Econometrica* 89, 1 (2021), 133–161.

[6] Elias Bareinboim and Judea Pearl. 2016. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences* 113, 27 (2016), 7345–7352.

[7] Alan D Berkowitz. 2005. An overview of the social norms approach. *Changing the culture of college drinking: A socially situated health communication campaign* 1 (2005), 193–214.

[8] Cristina Bicchieri. 2005. *The grammar of society: The nature and dynamics of social norms.* Cambridge University Press.

[9] Patrick Billingsley. 1999. *Convergence of Probability Measures.* Wiley.

[10] Board of Governors of the Federal Reserve System. 2023. 2022 Survey of Consumer Finances (SCF). https://www.federalreserve.gov/econres/scfindex.htm. doi:10.17016/8799 The Survey of Consumer Finances (SCF) is a triennial cross-sectional survey of U.S. families, providing data on balance sheets, pensions, income, and demographic characteristics..

[11] Adrienne Chung Adrienne Chung and Rajiv N Rimal Rajiv N Rimal. 2016. Social norms: A review. *Review of Communication Research* 4 (2016), 01–28.

[12] Gordon Dai, Weijia Zhang, Jinhan Li, Siqi Yang, Srihas Rao, Arthur Caetano, Misha Sra, et al. 2024. Artificial leviathan: Exploring social evolution of llm agents through the lens of hobbesian social contract theory. *arXiv preprint arXiv:2406.14373* (2024).

[13] Stefano DellaVigna and Devin Pope. 2018. What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies* 85, 2 (2018), 1029–1069.

[14] Miroslav Dudík, John Langford, and Lihong Li. 2011. Doubly robust policy evaluation and learning. *arXiv preprint arXiv:1103.4601* (2011).

[15] Gabriel Dulac-Arnold, Daniel Mankowitz, and Todd Hester. 2019. Challenges of real-world reinforcement learning. *arXiv preprint arXiv:1904.12901* (2019).

[16] Ernst Fehr and Simon Gächter. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90, 4 (2000), 980–994.

[17] Juan L Gamella and Christina Heinze-Deml. 2020. Active invariant causal prediction: Experiment selection through stability. *Advances in Neural Information Processing Systems* 33 (2020), 15464–15475.

[18] Ishaan Gulrajani and David Lopez-Paz. 2020. In search of lost domain generalization. *arXiv preprint arXiv:2007.01434* (2020).

[19] Robert XD Hawkins, Noah D Goodman, and Robert L Goldstone. 2019. The emergence of social norms and conventions. *Trends in cognitive sciences* 23, 2 (2019), 158–169.

[20] Michael Hechter and Karl-Dieter Opp. 2001. Social norms. (2001).

[21] Christina Heinze-Deml and Nicolai Meinshausen. 2021. Conditional variance penalties and domain shift robustness. *Machine Learning* 110, 2 (2021), 303–348.

[22] Miguel A Hernán and James M Robins. 2010. Causal inference.

[23] Guido W Imbens and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge university press.

[24] Nan Jiang and Lihong Li. 2016. Doubly robust off-policy value evaluation for reinforcement learning. In *International conference on machine learning*. PMLR, 652–661.

[25] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased learning-to-rank with biased feedback. In *Proceedings of the tenth ACM international conference on web search and data mining*. 781–789.

[26] Olaf Kallenberg. 2002. *Foundations of Modern Probability.* Springer.

[27] Erin L Krupka and Roberto A Weber. 2013. Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association* 11, 3 (2013), 495–524.

[28] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences* 116, 10 (2019), 4156–4165.

[29] David Lewis. 2008. *Convention: A philosophical study.* John Wiley & Sons.

[30] Chaochao Lu, Yuhuai Wu, José Miguel Hernández-Lobato, and Bernhard Schölkopf. 2021. Invariant causal representation learning for out-of-distribution generalization. In *International Conference on Learning Representations*.

[31] Sean P. Meyn and Richard L. Tweedie. 2009. *Markov Chains and Stochastic Stability.* Springer.

[32] Michael Oberst, Nikolaj Thams, Jonas Peters, and David Sontag. 2021. Regularizing towards causal invariance: Linear models with proxies. In *International Conference on Machine Learning*. PMLR, 8260–8270.

[33] Karl-Dieter Opp. 1982. The evolutionary emergence of norms. *British journal of social psychology* 21, 2 (1982), 139–149.

[34] Judea Pearl. 2009. *Causality.* Cambridge university press.

[35] Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. 2016. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 78, 5 (2016), 947–1012.

[36] Niklas Pfister, Peter Bühlmann, and Jonas Peters. 2019. Invariant causal prediction for sequential data. *J. Amer. Statist. Assoc.* 114, 527 (2019), 1264–1276.

[37] Paul R Rosenbaum and Donald B Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 1 (1983), 41–55.

[38] Dominik Rothenhäusler, Nicolai Meinshausen, Peter Bühlmann, and Jonas Peters. 2021. Anchor regression: Heterogeneous data meet causality. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83, 2 (2021), 215–246.

[39] Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence* 1, 5 (2019), 206–215.

[40] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*. PMLR, 1670–1679.

[41] Sandip Sen and Stéphane Airiau. 2007. Emergence of norms through social learning.. In *IJCAI*, Vol. 1507. 1512.

[42] Leigh S Shaffer. 1983. Toward Pepitone's vision of a normative social psychology: What is a social norm? *The journal of mind and behavior* (1983), 275–293.

[43] Yoav Shoham and Moshe Tennenholtz. 1995. On social laws for artificial agent societies: off-line design. *Artificial intelligence* 73, 1-2 (1995), 231–252.

[44] Yoav Shoham and Moshe Tennenholtz. 1997. On the emergence of social conventions: modeling, analysis, and simulations. *Artificial Intelligence* 94, 1-2 (1997), 139–166.

[45] Adith Swaminathan and Thorsten Joachims. 2015. Counterfactual risk minimization: Learning from logged bandit feedback. In *International conference on machine learning*. PMLR, 814–823.

[46] Philip Thomas and Emma Brunskill. 2016. Data-efficient off-policy policy evaluation for reinforcement learning. In *International conference on machine learning*. PMLR, 2139–2148.

[47] Jin Tian and Judea Pearl. 2000. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence* 28, 1 (2000), 287–313.

[48] Stefan Wager and Susan Athey. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* 113, 523 (2018), 1228–1242.

[49] Mengyue Yang, Zhen Fang, Yonggang Zhang, Yali Du, Furui Liu, Jean-Francois Ton, Jianhong Wang, and Jun Wang. 2023. Invariant learning via probability of sufficient and necessary causes. *Advances in Neural Information Processing Systems* 36 (2023), 79832–79857.

[50] Yuzhe Yang, Yifei Zhang, Minghao Wu, Kaidi Zhang, Yunmiao Zhang, Honghai Yu, Yan Hu, and Benyou Wang. 2025. TwinMarket: A Scalable Behavioral and Social Simulation for Financial Markets. In *ICLR 2025 Workshop on Advances in Financial AI.* arXiv:2502.01506.

[51] H Peyton Young. 1993. The evolution of conventions. *Econometrica: Journal of the Econometric Society* (1993), 57–84.

[52] H Peyton Young. 2015. The evolution of social norms. *Annual Review of Economics* 7, 1 (2015), 359–387.

[53] Kun Zhang, Mingming Gong, and Bernhard Schölkopf. 2015. Multi-source domain adaptation: A causal view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29.

A Assumptions and RL Plausibility

Compactness and Feller Continuity: Sufficient Primitives. Assume that the market evolution is governed by a continuous transition function $F^{(\theta)}$ with i.i.d. exogenous noise ξ_t. Specifically, if $X_{t+1} = F^{(\theta)}(X_t, \xi_t)$, and $F^{(\theta)}$ is continuous while ξ_t is i.i.d. with fixed law, then for any bounded continuous function $f \in C_b(\mathcal{X})$, the transition kernel $P^{(\theta)}$ defined by $P^{(\theta)} f(x) = \mathbb{E}[f(F^{(\theta)}(x, \xi))]$ is continuous by the dominated convergence theorem. Consequently, $P^{(\theta)}$ is Feller. Compactness of the state space can be enforced via clipping and bounded encoders.

Alternative Assumptions (used only if necessary). We could alternatively assume **weak-Feller + tightness** on Polish spaces, or **drift + minorization** (positive Harris recurrence). These provide drop-in replacements for existence and convergence theorems. However, the main text does not rely on these alternative sets of assumptions.

B Existence via Krylov–Bogolyubov

Standing assumptions: $(\mathcal{X}, d)$ is a compact metric space, and P is Feller on $(\mathcal{X}, \mathcal{B})$.

Lemma B.1 (Tightness of Cesàro averages). *Fix $x_0 \in \mathcal{X}$. Define the Cesàro averages of the Markov chain as $\nu_T \triangleq \frac{1}{T}\sum_{t=0}^{T-1} P^t(x_0, \cdot)$. Then the family $\{\nu_T\}_{T\geq 1}$ is tight in $\mathcal{P}(\mathcal{X})$, where $\mathcal{P}(\mathcal{X})$ denotes the space of probability measures on $\mathcal{X}$.*

Proof. Since $\mathcal{X}$ is compact, every probability measure on $\mathcal{X}$ is tight. Moreover, $\mathcal{P}(\mathcal{X})$ is compact in the weak topology, which guarantees the tightness of $\{\nu_T\}_{T\geq 1}$. □

Lemma B.2 (Invariance of weak limits). *If $\nu_{T_k} \Rightarrow \pi$ weakly, then π is invariant for P.*

Proof. Let $f \in C_b(\mathcal{X})$. We have

$$\int (Pf)\, d\nu_T - \int f\, d\nu_T = \frac{1}{T}\left(\int f\, dP^T(x_0, \cdot) - f(x_0)\right) \xrightarrow[T\to\infty]{} 0.$$

Since P is Feller, $Pf \in C_b(\mathcal{X})$. By weak convergence,

$$\int (Pf)\, d\nu_{T_k} \to \int (Pf)\, d\pi \quad \text{and} \quad \int f\, d\nu_{T_k} \to \int f\, d\pi.$$

Therefore, $\int (Pf)\, d\pi = \int f\, d\pi$ for all $f \in C_b(\mathcal{X})$, implying that $\pi P = \pi$. □

Theorem B.3 (Krylov–Bogolyubov Theorem). *The set of invariant measures $\mathcal{I}(P)$ is nonempty, convex, and weakly compact. In particular, for any $\phi \in C_b(\mathcal{X})$, the signature set*

$$\Sigma(P, \phi) = \left\{ \int \phi\, d\pi : \pi \in \mathcal{I}(P) \right\}$$

is nonempty and compact.

References. Billingsley [9], Kallenberg [26], Meyn and Tweedie [31].

C Occupation Measures and Time Averages

Standing assumptions: $(\mathcal{X}, d)$ is compact and P is Feller. The process $\{X_t\}_{t\geq 0}$ is a Markov chain with transition kernel P. Define $\mu_T \triangleq \frac{1}{T}\sum_{t=1}^{T} \delta_{X_t}$ as the occupation measure.

Lemma C.1 (Tightness of occupation measures). *The family* $\{\mu_T\}_{T\geq 1}$ *is tight in* $\mathcal{P}(\mathcal{X})$.

Lemma C.2 (Cesàro average of martingale differences). *For* $f \in C_b(\mathcal{X})$, *define* $M_{t+1} \triangleq f(X_{t+1}) - (Pf)(X_t)$ *as a bounded martingale difference. Then* $\frac{1}{T}\sum_{t=0}^{T-1} M_{t+1} \to 0$ *in* L^1 *and almost surely.*

Lemma C.3 (Invariance of weak limits of μ_T). *If* $\mu_{T_k} \Rightarrow \mu_\star$ *weakly, then* $\mu_\star \in \mathcal{I}(P)$.

Theorem C.4 (Limit points of time averages lie in Σ). *Let* $\phi \in C_b(\mathcal{X})$ *and* $\bar{\phi}_T = \int \phi\, d\mu_T$. *Every subsequential limit of* $(\bar{\phi}_T)$ *equals* $\int \phi\, d\pi$ *for some* $\pi \in \mathcal{I}(P)$. *Hence, all limit points belong to* $\Sigma(P,\phi)$.

If ϕ is only a bounded Borel function, weak convergence need not imply continuity of integrals. However, the conclusion of Theorem C.4 still holds if ϕ is π-almost surely continuous for every $\pi \in \mathcal{I}(P)$ or if there exist $(\phi_n) \subset C_b(\mathcal{X})$ such that $\|\phi_n - \phi\|_{L^1(\pi)} \to 0$ uniformly over $\pi \in \mathcal{I}(P)$.

D Convergence of Long-Run Statistics

The main text does *not* require convergence assumptions. For completeness we recall two sufficient conditions.

Definition D.1 (ϕ-uniqueness). P has *ϕ-uniqueness* if $\int \phi\, d\pi$ equals the same constant c_ϕ for all $\pi \in \mathcal{I}(P)$.

Theorem D.2 (ϕ-uniqueness $\Rightarrow$ convergence). *Under ϕ-uniqueness,* $\bar{\phi}_T \to c_\phi$ *almost surely.*

Harris ergodic LLN.. If the chain is ψ-irreducible, aperiodic, and positive Harris recurrent, there is a unique invariant probability π, and $\frac{1}{T}\sum_{t=1}^{T}\phi(X_t) \to \int \phi\, d\pi$ a.s. for all integrable ϕ [31, Thm. 17.0.1].

E Detailed Estimation Method and Confidence Interval Calculation

We estimate the Probability of Necessity and Sufficiency (PNS) using the following unbiased estimator:

$$\widehat{\mathrm{PNS}}_g(\theta \mid \psi) = \frac{1}{N_\psi} \sum_{e:\,\psi(X_e)=1} \mathbf{1}\big\{\, Y_{g,e}(\theta) = 1,\ Y_{g,e}(\theta_0) = 0 \big\}\,.$$

where N_ψ is the number of seeds satisfying $\psi(X_e) = 1$.

For confidence intervals of this estimator, we use the Wilson score interval for a binomial proportion. The confidence intervals can be computed as follows: Let $N = N_\psi, k = \sum_{e:\,\psi(X_e)=1} \mathbf{1}\{Y_{g,e}(\theta) = 1,\ Y_{g,e}(\theta_0) = 0\}$, and $\hat{p} = k/N$.

Wilson score interval:

$$\left(\frac{\hat{p} + \frac{z^2_{\alpha/2}}{2N} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N} + \frac{z^2_{\alpha/2}}{4N^2}}}{1 + \frac{z^2_{\alpha/2}}{N}},\quad \frac{\hat{p} + \frac{z^2_{\alpha/2}}{2N} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N} + \frac{z^2_{\alpha/2}}{4N^2}}}{1 + \frac{z^2_{\alpha/2}}{N}} \right).$$

F Experiment Setup

F.1 Social Norm Ranges

We instantiate per–group bands for ST (Subsidy/Transaction) and RI (Revenue/Investment). Group names follow the main text: **low-resource**, **mid-resource**, **high-resource**. Intervals use standard open/closed notation; ∞ denotes no upper bound.

ST-1 (low subsidy intensity across groups).

low-resource : $(0,\ 0.7]$,
mid-resource : $(0,\ 1.0]$,
high-resource : $(0,\ 1.3]$.

ST-2 (regressive subsidy: higher for low-resource, lower for high-resource).

low-resource : $[3.5,\ \infty)$,
mid-resource : $[2.3,\ \infty)$,
high-resource : $(0,\ 2.0]$.

RI-1 (upward mobility: higher RI for low-resource, lower for high-resource).

low-resource : $[1.05,\ \infty)$,
high-resource : $(0,\ 0.4]$.

RI-2 (concentration: lower RI for low-resource, higher for high-resource).

low-resource : $(0,\ 0.62]$,
high-resource : $(0.8,\ \infty)$.

These bands are used for the "norm attainment" event defined in the main text: a run attains a norm if each group's long-run statistic eventually stays within its band.

F.2 Initial Configurations

Key parameters for five economies (CRRA=risk aversion, IFE=effort elasticity, e_p=superstar prob., e_q=stability, ρ_e=persistence, σ_e=shock, $super_e$=multiplier), see Table 3. Experiments 1 and 3 use IC1, whereas the ablation study (Experiment 2) uses IC1–IC5.

Table 3: Initial-condition (IC) configurations used in all experiments.

IC	CRRA	IFE	e_p	e_q	ρ_e	σ_e	$super_e$
IC1	1.0	2.0	2.2×10^{-6}	0.990	0.982	0.20	504.3
IC2	1.5	1.8	5.0×10^{-6}	0.985	0.975	0.18	400
IC3	0.7	2.5	8.0×10^{-6}	0.985	0.990	0.15	350
IC4	1.2	2.2	1.0×10^{-5}	0.920	0.950	0.22	450
IC5	2.0	1.5	3.0×10^{-6}	0.995	0.990	0.25	600

F.3 Action Space: Platform and Users

To keep optimization well-posed under budget/resource constraints, we parameterize controls by *proportions* (ratios) rather than raw levels. This keeps the feasible set compact and stabilizes MARL training.

Table 4: Training parameters. One step = one year; evaluation window T=50 (post burn-in).

Item	Value
Population	$n_{\text{households}} = 100$
Episode length (env)	**300** steps (episode termination)
Model / Buffer	hidden size = 256; replay buffer = 10^6
Optimization	Adam; $lr = 3 \times 10^{-4}$; batch = 128; $\gamma_{\text{RL}} = 0.975$; $\tau = 5 \times 10^{-3}$
Schedule	1500 epochs; *epoch length* = 500 env steps (across episodes); update freq.=2; eval = 100 episodes
PPO (aux)	$\gamma_{\text{PPO}} = 0.99$; $\tau = 0.95$; clip = 0.1; $v_{\text{loss}} = 0.5$; ent= 0.01
Exploration	warm-up = 1000 steps; noise = 0.1; $\epsilon = 0.1$
Randomization	train seeds $\{0, \ldots, 7\}$; val $\{8, 9\}$; test $\{10, \ldots, 12\}$; twin-world PNS shares exogenous randomness for (θ, θ_0) per seed
Checkpointing	save model every 100 epochs

User actions. Each user i chooses an *investment share* $p_t^i \in (0, 1)$ (e.g., ads/promo budget ratio) and an *activity level* $h_t^i \in [0, h_{\max}]$ (e.g., listing/engagement frequency). Let rev_t^i denote realized revenue and inv_t^i the investment stock for user i. The platform applies a subsidy schedule $S(\cdot; \sigma_t, \kappa_t)$ and a fee schedule $F(\cdot; f_t, \kappa_{f_t})$.[3] Under the intertemporal budget,

$$p_t^i = \frac{inv_{t+1}^i}{rev_t^i + S(rev_t^i; \sigma_t, \kappa_t) - F(rev_t^i; f_t, \kappa_{f_t}) + inv_t^i}, \qquad h_t^i \in [0, h_{\max}],$$

so decision-making over (spend, effort) is implemented via proportional controls (p_t^i, h_t^i).

Platform actions. The platform sets high-level levers that shape incentives and aggregate allocation:

$$\mathcal{A}_t^{\text{P}} = \{\, \sigma_t,\ \kappa_t,\ f_t,\ \kappa_{f_t},\ \rho_t \,\}, \qquad R_t = \rho_t\, Y_t,$$

where σ_t and κ_t parameterize subsidies/exposure, f_t and κ_{f_t} parameterize fees/take-rate tiers, R_t is the platform's off–transaction spend, and Y_t is an aggregate proxy (e.g., total GMV).

Action spaces. With proportional parameterization and clipping,

$$\mathcal{A}_t^{u,i} = \{\, p_t^i \in (0, 1),\ h_t^i \in [0, h_{\max}] \,\}, \qquad \mathcal{A}_t^{\text{P}} \text{ as above.}$$

All actions are bounded to ensure compactness and continuity of the induced transition kernel.

G Training Parameters

Table 4 summarizes all training and optimization settings used in our experiments, including shared defaults and the ranges used in ablations. Unless otherwise noted, results use the Table 4 defaults.

[3]For example, tiered piecewise-linear schedules controlled by platform parameters; details omitted.

G.1 Seed Configurations

We use five independent contexts (IC1–IC5). Train and test seeds are strictly disjoint; at load time we drop any accidental overlaps.[4] Table 5 lists the exact seeds shipped with our logs; if a replication uses a different pool, the odd/even rule preserves disjointness.

Table 5: Seeds for ablation experiments. Train uses odd seeds; Test uses even seeds within each IC.

IC	Train seeds	Test seeds
IC1	1101, 1103, 1105, 1107	1102, 1104, 1106, 1108
IC2	1201, 1203, 1205, 1207	1202, 1204, 1206, 1208
IC3	1301, 1303, 1305, 1307	1302, 1304, 1306, 1308
IC4	1401, 1403, 1405, 1407	1402, 1404, 1406, 1408
IC5	1501, 1503, 1505, 1507	1502, 1504, 1506, 1508

G.2 Hyperparameters

Unless otherwise noted, we use the following defaults:

Stage I (PNS filter): $\tau_{\text{pns}} = 0.8$.

Stage II (routing objective): $\lambda_{\text{route}} = 0.1,\ \tau_{\text{cov}} = 0.2,\ \tau_{\text{prune}} = 10^{-2},\ K = 80,\ \lambda_{\text{len}} = 0.3,\ \epsilon_{\text{imp}} = 10^{-6}$.

Hybrid baseline weight: $\alpha = 0.7$ in $\alpha\,\overline{\text{Cov}} + (1 - \alpha)\,\overline{r}$.

Overall Perf weights: $w_{\text{pns}} = 0.8,\ w_{\text{cov}} = 0.2,\ w_{\text{gap}} = 0.1$.

Stage III (factor selection): $\theta_{\text{dist}} = 0.3,\ \alpha = 0.05$.

RNG seeds: $s = 42$.

Environment train/val/test seeds for IC1–IC5 follow Table 5 (odd for train, even for test).

Notes for Stage III. For factor f_k, distance on the complier set is the (normalized) 1-Wasserstein metric:

$$D_k = \frac{W_1(\widehat{F}_{k,\theta}, \widehat{F}_{k,\theta_0})}{s_k},$$

where s_k is a scale term.

A factor is declared *key* only if both hold: (i) $D_k \geq \theta_{\text{dist}}$ with default $\theta_{\text{dist}} = 0.3$, and (ii) a two-sided permutation test on W_1 is significant at $\alpha = 0.05$ after Holm–Bonferroni correction across all factor×group (and baseline, if multiple) tests within the route.

H Complexity and Overhead

Notation.

C : #(IC,norm,base,task), M : rows/CSV, G : groups,
R : rules (Stage I), B : buckets, U : eval items, L : final list len.

Stage II components.

$$\begin{aligned} \text{Target-set selection :} &\quad O(C\,G^2\,M) \\ \text{Scoring / filtering :} &\quad O(R\,C) + O(R \log R) \\ \text{Greedy routing :} &\quad O(L\,R\,B\,U) \\ \text{Backward pruning :} &\quad O(L^2\,B\,U) \end{aligned}$$

[4]Our runner removes from `test` any IC directory that also appears in `train`.

Table 6: Asymptotic complexity (Big-O).

Stage	Time	Space
Stage I (PNS)	$O(C\,M\,G)$	$O(1)$
Stage II (routing)	$O\big(CG^2M + RC + R\log R + LRBU + L^2BU\big)$	$O(RBU)$
Stage III (attribution)	$O(LU)$	$O(U)$

Overall. Dominant terms are typically $O(C\,G^2\,M)$ and $O(L\,R\,B\,U)$; Stage III is negligible relative to Stage II.