
Participation Dynamics in Two-Sided Platforms with AI Shadow Creators

Yurui Wu, Fan Yao

Department of Statistics & Operations Research, UNC Chapel Hill

Chapel Hill, NC, 27599

{yuruiwu, fanyao}@email.unc.edu

Abstract

Generative AI changes the supply side of content platforms by adding scalable non-human content production. This paper studies a fixed-policy participation model in which viewer populations respond to satisfaction, human creators respond to human exposure net of crowd-out, and AI shadow creators respond to AI exposure. The model extends two-sided platform participation dynamics by separating human and AI creator blocks and by allowing AI quality to depend on human data and on the diversity of the human creator ecology. We give primitive conditions for existence, uniqueness, global convergence, human dropout, and cyclic instability. In the two-field case, we derive certificates for the four possible human-support regimes: both fields active, only the first field active, only the second field active, and no human field active. These certificates are stated in terms of initial-domain and primitive quantities, not equilibrium objects. We also show that the relaxation parameter changes convergence and local stability, but not the fixed-point set. Synthetic experiments reproduce the theory: high human costs induce sharp dropout, human-share targets create an interior welfare peak, hot recommendation improves stage welfare while reducing ecological balance, and large relaxation can generate persistent period-two and period-four cycles. The results identify a mechanism by which AI supply can increase short-run satisfaction while endogenously eroding the human participation base on which long-run content quality may depend.

1 Introduction

Recommendation policies on a two-sided platform affect more than the matching realized in the current period. They also change the populations that return to the platform in the next period. Viewers may return when they receive satisfactory recommendations; creators may remain active when they obtain exposure. Recent dynamic models formalize this feedback loop and show that a policy with strong myopic performance can have weaker long-run consequences when participation is endogenous [11].

The introduction of generative AI makes the supply-side feedback loop more delicate. AI can scale content quickly and at low marginal cost. This can improve availability and short-run viewer satisfaction, especially when human supply is scarce. At the same time, AI content can crowd out human attention and reduce the exposure that keeps human creators active. If AI systems also learn from human-created data, then the platform faces a dynamic tradeoff: the supply source that helps satisfy current demand can weaken the human ecology that supports future quality. This paper builds a tractable participation model for this tradeoff.

We study a platform with viewer groups, human creator fields, and AI shadow creator fields. A fixed recommendation policy routes viewers to fields. Within each routed field, a preference-share parameter splits exposure between human and AI content. Viewer satisfaction depends on a field

quality aggregate and on a smooth human-AI mixture factor. Human creators receive a reference population from human exposure, but this reference is divided by an AI crowd-out term. AI shadow creators receive a reference population from AI exposure. The resulting dynamics are relaxed best-response updates in a finite-dimensional game.

The analysis has two purposes. First, it gives well-posed fixed-policy theory: existence, fixed-point/Nash equivalence, global convergence under a spectral comparison condition, and primitive dropout certificates. Second, it clarifies which comparative statics are about equilibria and which are about dynamics. Human costs, crowd-out coefficients, and target mixture shares can move equilibrium welfare and support. The relaxation parameter does not move the fixed-point set. It changes convergence speed and can destabilize a fixed point through a flip mode.

1.1 Main contributions

A human-AI participation model with primitive assumptions. We define a two-sided participation game with three state blocks: viewer populations, human creator populations, and AI shadow creator populations. All standing assumptions are imposed on primitive maps and on the compact domain generated by the initial condition. They do not assume properties of the unknown equilibrium. We explicitly define the derivative operator D_x and use local differentiability only inside local sensitivity theorems.

Existence, uniqueness, and convergence. For every fixed policy, the relaxed best-response map is continuous and maps a compact positive box into itself, so a steady-state equilibrium exists. Nash equilibria of the induced participation game coincide with best-response fixed points and with steady states of the dynamic. Under a computable nonnegative comparison matrix $M(\pi)$ with spectral radius less than one, the fixed point is unique and every trajectory converges geometrically under a weighted block norm.

Dropout certificates for two fields. For $K = L = 2$, we give primitive sufficient conditions for all four human-support regimes: both human fields active, only field 1 active, only field 2 active, and no human field active. The certificates use lower exposure bounds generated by the primitive viewer floor and upper reference bounds generated by the compact domain. The same framework yields explicit intervals for a human-cost multiplier and for a field-level crowd-out coefficient. Quantities that are automatically nonnegative are not introduced as positivity assumptions; strict positivity is required only when a finite crowd-out threshold is desired.

Dynamic role of the relaxation parameter. We prove that the common relaxation parameter η does not change the equilibrium set. When the primitive best-response comparison matrix is a contraction, larger η improves the comparison-rate bound. Locally, however, a negative eigenmode can cross -1 as η increases, generating persistent oscillations. The simulations reproduce this distinction: welfare is nearly invariant across stable η values but collapses into a period-two regime when η is too large.

Simulation evidence and diagnostic figures. We run the supplied simulation code and incorporate the resulting figures and tables. The experiments show stable convergence in a baseline environment, a policy tradeoff between myopic welfare and ecology, sharp human dropout under increasing human cost, an interior welfare peak in the target human share, and nonconvergent period-two/period-four dynamics under aggressive relaxation and steep reference maps.

2 Related work

Two-sided platforms and participation dynamics. The paper is related to the economics of two-sided platforms, where platform value depends on interactions between distinct user populations [1, 14, 15]. Our focus is narrower: we fix a recommendation policy and study the induced participation dynamics. The closest dynamic benchmark is Kiyohara et al. [11], who study policy design for two-sided platforms with viewer and creator participation. We extend that participation framework by splitting the creator side into human and AI blocks and by adding crowd-out and ecology channels.

Strategic creators and recommender systems. A growing literature studies content providers as strategic or adaptive agents. Ben-Porat and Tennenholtz [2] formulate recommendation with strategic

content providers. Hron et al. [5] model creator incentives on algorithm-curated platforms. Yao et al. [18] analyze top- K recommendation under competing content creators, and Dai et al. [3] study incentive-aware recommender systems. Huttenlocher et al. [7] examine matching users and creators when departures are possible. Our model differs by making human departure interact with AI supply through an explicit population state.

Exposure, fairness, and ecosystem objectives. Exposure allocation is central in ranking and recommendation. Singh and Joachims [17] introduce exposure fairness in rankings, and Patro et al. [12] study two-sided fairness in personalized recommendation. Our welfare criterion is viewer-side stage satisfaction, but exposure matters because it controls future human participation. The ecology channel in our model is not a fairness constraint; it is a smooth, diversity-based mechanism through which a balanced and active human creator ecology can improve future AI quality. Its diversity component follows the Shannon diversity index used in information theory and ecology, while the activity multiplier makes the score vanish when the human creator base is economically inactive [4, 9, 13, 16].

Feedback loops and generative AI. Feedback loops in recommender systems can amplify endogenous shifts in the population state [8]. Generative AI introduces a new supply-side source in such loops. Recent work studies generative AI in user-generated content environments [6], platform welfare effects of AI-generated supply [19], and empirical evidence on whether generative AI crowds out human creators [10]. We provide a stylized dynamic theory for one mechanism behind such crowd-out: AI exposure reduces the human reference population, and the resulting human decline feeds back into satisfaction and AI quality.

3 Problem setup

3.1 Platform, policy, and state

Let $\mathcal{K} = \{1, \dots, K\}$ be viewer groups and $\mathcal{L} = \{1, \dots, L\}$ be creator fields. Time is discrete. The platform policy is a row-stochastic routing matrix

$$\Pi = \left\{ \pi \in \mathbb{R}_+^{K \times L} : \sum_{\ell=1}^L \pi_{k\ell} = 1 \text{ for every } k \right\}. \quad (1)$$

The state is

$$x_t = (v_t, h_t, \theta_t) \in \mathbb{R}_+^{K+2L}, \quad (2)$$

where $v_t \in \mathbb{R}_+^K$ is the viewer population, $h_t \in \mathbb{R}_+^L$ is the human creator population, and $\theta_t \in \mathbb{R}_+^L$ is the AI shadow creator population.

Inside a routed field, exposure is split by primitive human-share parameters $\varrho_{k\ell} \in [0, 1]$. Under policy π , field ℓ receives human and AI exposure

$$e_\ell^H(v; \pi) = \sum_{k=1}^K \varrho_{k\ell} \pi_{k\ell} v_k, \quad (3)$$

$$e_\ell^A(v; \pi) = \sum_{k=1}^K (1 - \varrho_{k\ell}) \pi_{k\ell} v_k. \quad (4)$$

The total exposure in field ℓ is $e_\ell^H + e_\ell^A = \sum_k \pi_{k\ell} v_k$.

The smoothed human share in field ℓ is

$$r_\ell(h, \theta) = \frac{h_\ell + \epsilon_r}{h_\ell + \theta_\ell + 2\epsilon_r}, \quad \epsilon_r > 0, \quad (5)$$

so $r_\ell \in (0, 1)$ on the nonnegative orthant. Given target share $\omega_\ell \in (0, 1)$ and sensitivity $\kappa_\ell > 0$, define

$$m_\ell(r) = \exp\{-\kappa_\ell(r - \omega_\ell)^2\}. \quad (6)$$

This term is maximized when the realized human share matches the target.

3.2 Quality, satisfaction, and reference populations

Human quality in field ℓ for viewer group k is $f_{k\ell}^H(h_\ell)$. Intrinsic AI quality is $\tilde{f}_{k\ell}^A(\theta_\ell)$. AI quality also depends on human data in the same field and on a global human ecology score. Let

$$d_\ell(h_\ell) = \log(1 + h_\ell), \quad (7)$$

$$p_j(h) = \frac{h_j + \epsilon_E}{\sum_{i=1}^L (h_i + \epsilon_E)}, \quad \mathcal{D}_{\epsilon_E}(h) = - \sum_{j=1}^L p_j(h) \log p_j(h), \quad (8)$$

$$a_{\epsilon_E}(h) = \frac{\sum_{j=1}^L h_j}{\sum_{j=1}^L h_j + L\epsilon_E}, \quad \mathcal{E}_{\epsilon_E}(h) = a_{\epsilon_E}(h) \mathcal{D}_{\epsilon_E}(h), \quad (9)$$

with $\epsilon_E > 0$. The component $\mathcal{D}_{\epsilon_E}$ is the smoothed Shannon diversity of the human distribution across fields. Shannon diversity is a standard basis for ecological diversity and evenness diagnostics; the smoothing avoids boundary singularities. The multiplier a_{ϵ_E} is an activity correction: it drives $\mathcal{E}_{\epsilon_E}$ to zero when total human creator activity is zero, so a formally uniform but inactive creator distribution is not counted as a healthy ecology. Thus $\mathcal{E}_{\epsilon_E}$ is the reported ecology quantity in the simulations and the object that enters AI quality; raw diversity is not treated as a separate experimental outcome.

The full AI quality contribution is

$$f_{k\ell}^A(\theta_\ell; h_\ell, h) = \tilde{f}_{k\ell}^A(\theta_\ell) + \gamma_{k\ell}^{\text{data}} d_\ell(h_\ell) + \gamma^{\text{ent}} \mathcal{E}_{\epsilon_E}(h). \quad (10)$$

Define the field quality aggregate

$$Q_{k\ell}(h, \theta) = b_{k\ell} + \varrho_{k\ell} f_{k\ell}^H(h_\ell) + (1 - \varrho_{k\ell}) f_{k\ell}^A(\theta_\ell; h_\ell, h). \quad (11)$$

The satisfaction of viewer group k is

$$s_k(h, \theta; \pi) = \sum_{\ell=1}^L \pi_{k\ell} m_\ell(r_\ell(h, \theta)) Q_{k\ell}(h, \theta). \quad (12)$$

Each viewer group has a reference-population map $\bar{\lambda}_k : \mathbb{R} \rightarrow \mathbb{R}_+$. Each human field has $\bar{h}_\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, and each AI field has $\bar{\theta}_\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. Human participation is reduced by human cost $c_\ell \geq 0$ and crowd-out $\xi_\ell \geq 0$.

For fixed π , define the best responses

$$\text{BR}_k^V(x; \pi) = \bar{\lambda}_k(s_k(h, \theta; \pi)), \quad (13)$$

$$\text{BR}_\ell^H(x; \pi) = \left[\frac{\bar{h}_\ell(e_\ell^H(v; \pi))}{1 + \xi_\ell \theta_\ell} - c_\ell \right]_+, \quad (14)$$

$$\text{BR}_\ell^A(x; \pi) = \bar{\theta}_\ell(e_\ell^A(v; \pi)), \quad (15)$$

where $[z]_+ = \max\{z, 0\}$. With relaxation parameters $\eta_V, \eta_H, \eta_A \in (0, 1]$, the dynamics are

$$x_{t+1} = g(x_t, \pi) = x_t + H \{\text{BR}(x_t; \pi) - x_t\}, \quad (16)$$

where

$$H = \text{diag}(\underbrace{\eta_V, \dots, \eta_V}_K, \underbrace{\eta_H, \dots, \eta_H}_L, \underbrace{\eta_A, \dots, \eta_A}_L). \quad (17)$$

3.3 Standing primitive assumptions

The next assumptions are imposed on primitive maps and on the state domain. They do not refer to an unknown equilibrium.

Assumption 3.1 (Primitive boundedness and continuity). *There exist finite constants $V_{\max}, H_{\max}, A_{\max} > 0$ such that*

$$0 \leq \bar{\lambda}_k(s) \leq V_{\max}, \quad 0 \leq \bar{h}_\ell(e) \leq H_{\max}, \quad 0 \leq \bar{\theta}_\ell(e) \leq A_{\max} \quad (18)$$

for all k, ℓ and all arguments in their domains. The maps $\bar{\lambda}_k, \bar{h}_\ell, \bar{\theta}_\ell, f_{k\ell}^H, \tilde{f}_{k\ell}^A$ are continuous on their domains, and $\bar{h}_\ell, \bar{\theta}_\ell$ are nondecreasing. The smoothing constants satisfy $\epsilon_r, \epsilon_E > 0$.

The compact state domain is

$$\mathcal{X} = [0, V_{\max}]^K \times [0, H_{\max}]^L \times [0, A_{\max}]^L. \quad (19)$$

All fixed-policy results below are stated for initial conditions $x_0 \in \mathcal{X}$. This is not restrictive for bounded reference maps: after one update the best-response component is in $\mathcal{X}$, and convex relaxation keeps the trajectory in $\mathcal{X}$ whenever it starts in $\mathcal{X}$.

Assumption 3.2 (Nonnegative satisfaction primitives). *For all k, ℓ and nonnegative arguments, $b_{k\ell} \geq 0$, $f_{k\ell}^H \geq 0$, $\tilde{f}_{k\ell}^A \geq 0$, $\gamma_{k\ell}^{\text{data}} \geq 0$, and $\gamma^{\text{ent}} \geq 0$.*

Assumption 3.2 is used only when we claim nonnegativity of satisfaction and welfare. Existence and convergence do not require it.

Let

$$\begin{aligned} Q_{\max} = & \max_{k\ell} |b_{k\ell}| + \max_{k\ell} \sup_{h \in [0, H_{\max}]} |f_{k\ell}^H(h)| + \max_{k\ell} \sup_{a \in [0, A_{\max}]} |\tilde{f}_{k\ell}^A(a)| \\ & + \max_{k\ell} \gamma_{k\ell}^{\text{data}} \log(1 + H_{\max}) + \gamma^{\text{ent}} \log L. \end{aligned} \quad (20)$$

Since $m_\ell \leq 1$ and the policy rows sum to one, every satisfaction coordinate on $\mathcal{X}$ lies in $[-Q_{\max}, Q_{\max}]$. Also set $E_{\max} = KV_{\max}$.

Assumption 3.3 (Lipschitz primitives on the relevant compact sets). *The maps $f_{k\ell}^H$ are Lipschitz on $[0, H_{\max}]$, the maps $\tilde{f}_{k\ell}^A$ are Lipschitz on $[0, A_{\max}]$, the maps $\bar{\lambda}_k$ are Lipschitz on $[-Q_{\max}, Q_{\max}]$, and the maps $\bar{h}_\ell, \bar{\theta}_\ell$ are Lipschitz on $[0, E_{\max}]$. Let the maximum Lipschitz constants be denoted by $L_{fH}, L_{fA}, L_\lambda, L_h$, and L_θ .*

Definition 3.4 (Derivative notation). *For a map $F : \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}^m$ that is differentiable at (x, τ) , $D_x F(x, \tau)$ denotes the Jacobian of F with respect to the state variable x , regarded as a linear map from $\mathbb{R}^d$ to $\mathbb{R}^m$. The notation $\partial_\tau F(x, \tau)$ denotes the derivative with respect to the scalar or vector parameter τ with x held fixed.*

3.4 Induced game and welfare

For fixed π , the best responses arise from a simple concave game. Viewer group k chooses $v_k \geq 0$ and has utility

$$u_k(x; \pi) = v_k \bar{\lambda}_k(s_k(h, \theta; \pi)) - \frac{v_k^2}{2}. \quad (21)$$

Human field ℓ chooses $h_\ell \geq 0$ and has utility

$$u_\ell^H(x; \pi) = h_\ell \left(\frac{\bar{h}_\ell(e_\ell^H(v; \pi))}{1 + \xi_\ell \theta_\ell} - c_\ell \right) - \frac{h_\ell^2}{2}. \quad (22)$$

AI field ℓ chooses $\theta_\ell \geq 0$ and has utility

$$u_\ell^A(x; \pi) = \theta_\ell \bar{\theta}_\ell(e_\ell^A(v; \pi)) - \frac{\theta_\ell^2}{2}. \quad (23)$$

Each utility is strictly concave in its own scalar action, giving the best responses in (13)-(15).

The stage welfare criterion is viewer-side satisfaction

$$W(\pi; x) = \sum_{k=1}^K v_k s_k(h, \theta; \pi). \quad (24)$$

Proposition 3.5 (Well-defined welfare and nonnegativity). *Under Assumption 3.1, $W(\pi; x)$ is well defined for every $(\pi, x) \in \Pi \times \mathcal{X}$. Under Assumption 3.2, $W(\pi; x) \geq 0$.*

4 Fixed-policy equilibrium and convergence

4.1 Equilibrium, steady state, and dropout

Proposition 4.1 (Nash equilibria, best-response fixed points, and steady states). *Fix $\pi \in \Pi$. A state $x^* \in \mathcal{X}$ is a Nash equilibrium of the induced participation game if and only if $x^* = \text{BR}(x^*; \pi)$. It is a steady state of (16) if and only if $x^* = \text{BR}(x^*; \pi)$. Hence the three notions coincide.*

Theorem 4.2 (Existence and forward invariance). *Under Assumption 3.1, for every fixed policy $\pi \in \Pi$, the map $x \mapsto g(x, \pi)$ has at least one fixed point in $\mathcal{X}$. Moreover, if $x_0 \in \mathcal{X}$, then the trajectory generated by (16) remains in $\mathcal{X}$ for all $t \geq 0$.*

Proposition 4.3 (Field-wise human dropout). *Fix $\pi \in \Pi$ and $x \in \mathcal{X}$. For field ℓ ,*

$$\text{BR}_\ell^H(x; \pi) = 0 \iff \bar{h}_\ell(e_\ell^H(v; \pi)) \leq c_\ell(1 + \xi_\ell \theta_\ell). \quad (25)$$

In particular, larger human cost and larger AI crowd-out both weakly enlarge the dropout region.

4.2 A comparison matrix for global convergence

Let $A^H(\pi), A^A(\pi) \in \mathbb{R}_+^{L \times K}$ be the exposure matrices

$$A_{\ell k}^H(\pi) = \varrho_{k\ell} \pi_{k\ell}, \quad A_{\ell k}^A(\pi) = (1 - \varrho_{k\ell}) \pi_{k\ell}. \quad (26)$$

Then $e^H(v; \pi) = A^H(\pi)v$ and $e^A(v; \pi) = A^A(\pi)v$.

Lemma 4.4 (Block Lipschitz bounds). *Under Assumptions 3.1 and 3.3, for every fixed $\pi \in \Pi$ there exist finite constants $L_{s,H}$ and $L_{s,A}$ such that for all $x = (v, h, \theta)$ and $x' = (v', h', \theta')$ in $\mathcal{X}$,*

$$\|s(h, \theta; \pi) - s(h', \theta'; \pi)\| \leq L_{s,H} \|h - h'\| + L_{s,A} \|\theta - \theta'\|, \quad (27)$$

$$\|e^H(v; \pi) - e^H(v'; \pi)\| \leq \|A^H(\pi)\|_2 \|v - v'\|, \quad (28)$$

$$\|e^A(v; \pi) - e^A(v'; \pi)\| \leq \|A^A(\pi)\|_2 \|v - v'\|. \quad (29)$$

Moreover, with $\Xi_{\max} = \max_\ell \xi_\ell$,

$$\left| \frac{\bar{h}_\ell(e)}{1 + \xi_\ell \theta} - \frac{\bar{h}_\ell(e')}{1 + \xi_\ell \theta'} \right| \leq L_h |e - e'| + \Xi_{\max} H_{\max} |\theta - \theta'|. \quad (30)$$

For two states $x = (v, h, \theta)$ and $x' = (v', h', \theta')$, define the block-distance vector

$$\Delta(x, x') = \begin{bmatrix} \|v - v'\| \\ \|h - h'\| \\ \|\theta - \theta'\| \end{bmatrix} \in \mathbb{R}_+^3. \quad (31)$$

The comparison matrix is

$$M(\pi) = \begin{bmatrix} 1 - \eta_V & \eta_V L_\lambda L_{s,H} & \eta_V L_\lambda L_{s,A} \\ \eta_H L_h \|A^H(\pi)\|_2 & 1 - \eta_H & \eta_H \Xi_{\max} H_{\max} \\ \eta_A L_\theta \|A^A(\pi)\|_2 & 0 & 1 - \eta_A \end{bmatrix}. \quad (32)$$

Lemma 4.5 (Comparison inequality). *For all $x, x' \in \mathcal{X}$,*

$$\Delta(g(x, \pi), g(x', \pi)) \leq M(\pi) \Delta(x, x') \quad (33)$$

componentwise.

Theorem 4.6 (Unique steady state and global geometric convergence). *Fix $\pi \in \Pi$ and suppose $\rho(M(\pi)) < 1$, where $\rho(\cdot)$ denotes spectral radius. Then the fixed-policy dynamic has a unique steady-state equilibrium $x^*(\pi) \in \mathcal{X}$. For every $x_0 \in \mathcal{X}$, the trajectory under $\pi_t \equiv \pi$ converges to $x^*(\pi)$ geometrically under a weighted block-max norm. In particular, no nontrivial periodic orbit exists in $\mathcal{X}$ under this condition.*

Simulation counterpart. The baseline trajectory in Figure 1 is the direct numerical analogue of the fixed-policy theory: the three population blocks settle to a steady state, welfare and ecology stabilize, and the best-response residuals vanish. The limiting residual span is 2.16×10^{-15} in Table 2, so the numerical endpoint satisfies the fixed-point equations to machine precision. This does not prove the spectral condition, but it checks the qualitative prediction of Theorem 4.6 in the baseline parameterization.

5 Two-field comparative statics and support regimes

This section specializes to $K = L = 2$. The aim is to keep three issues separate: support regimes, local equilibrium sensitivity, and the dynamic role of relaxation. We write $x = (v_1, v_2, h_1, h_2, \theta_1, \theta_2)$.

5.1 Primitive support certificates

Define the primitive viewer floor

$$\underline{v}_k = \inf_{s \in [-Q_{\max}, Q_{\max}]} \bar{\lambda}_k(s), \quad k = 1, 2. \quad (34)$$

Because $\bar{\lambda}_k$ maps into $\mathbb{R}_+$, these floors are automatically nonnegative. No additional nonnegativity assumption is needed. The lower exposure bounds induced by policy π are

$$\underline{e}_\ell^H(\pi) = \sum_{k=1}^2 \varrho_{k\ell} \pi_{k\ell} \underline{v}_k, \quad \underline{e}_\ell^A(\pi) = \sum_{k=1}^2 (1 - \varrho_{k\ell}) \pi_{k\ell} \underline{v}_k. \quad (35)$$

They are also automatically nonnegative. Let

$$U_\ell = \sup_{e \in [0, E_{\max}]} \bar{h}_\ell(e), \quad A_\ell(\pi) = \frac{\bar{h}_\ell(\underline{e}_\ell^H(\pi))}{1 + \xi_\ell A_{\max}}. \quad (36)$$

The quantity $A_\ell(\pi)$ is a primitive survival certificate: it uses the worst allowed AI mass $A_{\max}$ and the lower human-exposure bound. The quantity U_ℓ is a primitive upper bound on the human reference population.

Theorem 5.1 (Two-field human-support certificates). *Fix a policy π and a subset $S \subseteq \{1, 2\}$. If*

$$c_\ell < A_\ell(\pi) \quad \text{for every } \ell \in S, \quad c_\ell \geq U_\ell \quad \text{for every } \ell \notin S, \quad (37)$$

then every fixed-policy equilibrium has human support S :

$$h_\ell^* > 0 \text{ for } \ell \in S, \quad h_\ell^* = 0 \text{ for } \ell \notin S. \quad (38)$$

Consequently, the four regimes have the following primitive certificates:

Regime	Sufficient primitive conditions
Both human fields active	$c_1 < A_1(\pi)$ and $c_2 < A_2(\pi)$
Only field 1 active	$c_1 < A_1(\pi)$ and $c_2 \geq U_2$
Only field 2 active	$c_2 < A_2(\pi)$ and $c_1 \geq U_1$
No human field active	$c_1 \geq U_1$ and $c_2 \geq U_2$

The certificates are sufficient rather than necessary. The undecided band $A_\ell(\pi) \leq c_\ell < U_\ell$ is precisely where the primitive lower and upper bounds do not determine the field's support for all equilibria.

5.2 Human-cost multiplier

Let $c_\ell(\alpha) = \alpha c_\ell^0$ with $\alpha > 0$ and $c_\ell^0 > 0$. The preceding theorem immediately gives explicit intervals.

Corollary 5.2 (Cost-scale intervals for the four support regimes). *Fix π . Define $a_\ell = A_\ell(\pi)/c_\ell^0$ and $u_\ell = U_\ell/c_\ell^0$. Then every equilibrium has:*

$$\text{both fields active} \quad \text{if } \alpha < \min\{a_1, a_2\}, \quad (39)$$

$$\text{only field 1 active} \quad \text{if } u_2 \leq \alpha < a_1, \quad (40)$$

$$\text{only field 2 active} \quad \text{if } u_1 \leq \alpha < a_2, \quad (41)$$

$$\text{no human field active} \quad \text{if } \alpha \geq \max\{u_1, u_2\}. \quad (42)$$

The one-field intervals are nonempty only when their left endpoint is smaller than their right endpoint.

Simulation counterpart. The cost sweep in Figure 3 implements the same threshold mechanism in the four-field synthetic environment. Human mass falls from 2.589 at zero cost to 0.404 at cost multiplier 2.0; at that point at least one field has already dropped out because the minimum field-level human population is numerically zero, while the maximum field-level human population remains 0.217. At multiplier 3.0 and above, all human fields are numerically extinct and welfare stays at the AI-only plateau of 1.812; see Table 5. The transition is sharper than the primitive sufficient certificates because those certificates are uniform over all equilibria in the compact domain.

5.3 Crowd-out coefficient in one field

Now fix all primitives except ξ_j for one field $j \in \{1, 2\}$. Let i denote the other field. Define

$$B_j(\pi) = \bar{h}_j(\underline{e}_j^H(\pi)), \quad \underline{\theta}_j(\pi) = \bar{\theta}_j(\underline{e}_j^A(\pi)). \quad (43)$$

The floor $\underline{\theta}_j(\pi)$ is nonnegative by construction. A strictly positive floor is needed only to obtain a finite off-threshold in ξ_j .

Assume $c_j > 0$ and define the active and inactive certificates

$$\xi_j^{\text{on}}(\pi) = \frac{1}{A_{\max}} \left(\frac{B_j(\pi)}{c_j} - 1 \right), \quad (44)$$

$$\xi_j^{\text{off}}(\pi) = \begin{cases} \max \left\{ 0, \frac{U_j/c_j - 1}{\underline{\theta}_j(\pi)} \right\}, & \underline{\theta}_j(\pi) > 0, \\ 0, & \underline{\theta}_j(\pi) = 0 \text{ and } c_j \geq U_j, \\ +\infty, & \underline{\theta}_j(\pi) = 0 \text{ and } c_j < U_j. \end{cases} \quad (45)$$

When $\xi_j^{\text{on}} \leq 0$, the primitive active certificate is empty.

Theorem 5.3 (Crowd-out thresholds and two-field support). *Fix π and vary only ξ_j .*

1. *If $0 \leq \xi_j < \xi_j^{\text{on}}(\pi)$, then field j is active at every equilibrium.*
2. *If $\xi_j \geq \xi_j^{\text{off}}(\pi)$, then field j is inactive at every equilibrium.*

Combining these two statements with the primitive status of field i gives all four two-field support regimes. For example, if $c_i < A_i(\pi)$ and $\xi_j \geq \xi_j^{\text{off}}(\pi)$, then only field i is active; if $c_i \geq U_i$ and $\xi_j \geq \xi_j^{\text{off}}(\pi)$, then no human field is active.

5.4 Local target-share sensitivity

Let τ be a scalar parameter and write $F(x, \tau) = \text{BR}(x; \pi, \tau) - x$. A local sensitivity formula requires differentiability and an invertible Jacobian at the equilibrium under consideration. This is not a standing assumption on the model; it is a local regularity condition for the following calculation.

Theorem 5.4 (Local equilibrium sensitivity). *Suppose x_0 satisfies $F(x_0, \tau_0) = 0$. For each human coordinate, assume that the pre-projection term in (14) has a strict local sign near (x_0, τ_0) : either it is strictly positive in a neighborhood, so $[\cdot]_+$ is locally the identity, or it is strictly negative in a neighborhood, so $[\cdot]_+$ is locally the zero map. Assume also that all primitive maps entering BR are C^1 in that neighborhood and that $I_6 - D_x \text{BR}(x_0; \pi, \tau_0)$ is invertible. Then there is a unique C^1 equilibrium branch $\tau \mapsto x^*(\tau)$ near τ_0 with $x^*(\tau_0) = x_0$, and*

$$\frac{dx^*}{d\tau} = (I_6 - D_x \text{BR}(x^*(\tau); \pi, \tau))^{-1} \partial_\tau \text{BR}(x^*(\tau); \pi, \tau). \quad (46)$$

The reduced welfare $\mathcal{W}(\tau) = W(\pi; x^*(\tau), \tau)$ satisfies

$$\frac{d\mathcal{W}}{d\tau} = D_x W(\pi; x^*(\tau), \tau) \frac{dx^*}{d\tau} + \partial_\tau W(\pi; x^*(\tau), \tau). \quad (47)$$

Consider a common target share $\omega_1 = \omega_2 = \omega$. Let

$$B_\ell(x) = \sum_{k=1}^2 v_k \pi_{k\ell} Q_{k\ell}(h, \theta), \quad \ell = 1, 2. \quad (48)$$

Under Assumption 3.2, $B_\ell(x) \geq 0$.

Proposition 5.5 (Target-share derivative). *Under the conditions of Theorem 5.4 with $\tau = \omega$,*

$$\frac{d\mathcal{W}}{d\omega} = \sum_{\ell=1}^2 2\kappa_\ell(r_\ell^* - \omega) m_\ell(r_\ell^*) B_\ell(x^*) + D_x W(I_6 - D_x \text{BR})^{-1} \partial_\omega \text{BR}, \quad (49)$$

where all quantities on the right are evaluated at $(x^*(\omega), \omega)$. If $r_1^* = r_2^* = \omega$ at some regular equilibrium, then $\partial_\omega \text{BR} = 0$, $dx^*/d\omega = 0$, and $d\mathcal{W}/d\omega = 0$ at that point.

The direct term in (49) is positive when the realized share is above target and negative when it is below target. Thus, if the equilibrium human share is around 0.39, increasing a common target far beyond that value depresses the mixture factor and can reduce welfare.

Simulation counterpart. Figure 4 shows this derivative logic in the common-target sweep. Welfare peaks at $\omega = 0.40$, where the realized mean human share is 0.391 and the mean mixture factor is 0.977. When the target is raised to 0.95, the realized share remains only 0.376, the target-realization gap grows to 0.574, the mean mixture factor falls to 0.374, and welfare drops to 1.195.

5.5 Relaxation parameter

Suppose all blocks use a common relaxation parameter $\eta \in (0, 1]$:

$$g_\eta(x; \pi) = (1 - \eta)x + \eta \text{BR}(x; \pi). \quad (50)$$

Proposition 5.6 (Relaxation does not move the fixed-point set). *For every $\eta \in (0, 1]$,*

$$g_\eta(x; \pi) = x \iff \text{BR}(x; \pi) = x. \quad (51)$$

Therefore η does not create a new equilibrium and does not move an existing equilibrium. It changes the path and stability of the relaxed dynamic.

In the common- η case, (32) takes the form

$$M_\eta(\pi) = (1 - \eta)I_3 + \eta C(\pi), \quad (52)$$

where

$$C(\pi) = \begin{bmatrix} 0 & L_\lambda L_{s,H} & L_\lambda L_{s,A} \\ L_h \|A^H(\pi)\|_2 & 0 & \Xi_{\max} H_{\max} \\ L_\theta \|A^A(\pi)\|_2 & 0 & 0 \end{bmatrix}. \quad (53)$$

Theorem 5.7 (Global rate effect of common relaxation). *If $\rho(C(\pi)) < 1$, then for every $\eta \in (0, 1]$ the fixed-policy dynamic has the same unique steady-state equilibrium and converges globally to it. Moreover,*

$$\rho(M_\eta(\pi)) = 1 - \eta\{1 - \rho(C(\pi))\}, \quad (54)$$

which is strictly decreasing in η .

The preceding theorem is a global comparison result. Local stability can still fail outside the contraction region.

Theorem 5.8 (Local stability threshold). *Suppose $\text{BR}(\cdot; \pi)$ is C^1 near a fixed point x^* , and let $J = D_x \text{BR}(x^*; \pi)$. Then*

$$D_x g_\eta(x^*; \pi) = (1 - \eta)I + \eta J. \quad (55)$$

The linearized dynamic is Schur stable if and only if $|1 - \eta + \eta\nu| < 1$ for every eigenvalue ν of J ; in that case x^* is locally asymptotically stable for the nonlinear relaxed dynamic. If some multiplier has modulus larger than one, the fixed point is linearly unstable. For a real eigenvalue $\nu < -1$, the corresponding multiplier crosses -1 at

$$\eta_{\text{flip}} = \frac{2}{1 - \nu} \in (0, 1), \quad (56)$$

so increasing η past this threshold destabilizes that linear mode.

Simulation counterpart. The relaxation sweep in Figure 2 separates fixed-point location from dynamic stability. For $\eta \in [0.10, 0.90]$, the limiting viewer, human, AI, and welfare values agree up to numerical precision. At $\eta = 0.95$ and $\eta = 1.00$, the same experimental family enters a period-two regime, with tail spans 0.914 and 1.123, respectively. The period-two aggregate cycle alternates between $(v, h, \theta) = (0.424, 1.597, 1.408)$ and $(2.880, 0, 0.465)$; see Table 4. The high-slope sweep in Figure 2 further shows persistent period-four behavior in the high-relaxation, high-slope region, matching the local flip-mode interpretation of Theorem 5.8.

6 One-field benchmark

The case $K = L = 1$ removes cross-field notation and exposes the basic coexistence mechanism. Let v be the viewer population, h the human population, θ the AI population, and $\varrho \in [0, 1]$ the human

exposure share. Since $L = 1$, the cross-field diversity component is zero. The equations are

$$e^H = \varrho v, \quad e^A = (1 - \varrho)v, \quad (57)$$

$$r(h, \theta) = \frac{h + \epsilon_r}{h + \theta + 2\epsilon_r}, \quad m(r) = \exp\{-\kappa(r - \omega)^2\}, \quad (58)$$

$$q(h, \theta) = b + \varrho f^H(h) + (1 - \varrho)\{\tilde{f}^A(\theta) + \gamma^{\text{data}} \log(1 + h)\}, \quad (59)$$

$$s(h, \theta) = m(r(h, \theta))q(h, \theta). \quad (60)$$

The equilibrium system is

$$v = \bar{\lambda}(s(h, \theta)), \quad (61)$$

$$h = \left[\frac{\bar{h}(\varrho v)}{1 + \xi \theta} - c \right]_+, \quad (62)$$

$$\theta = \bar{\theta}((1 - \varrho)v). \quad (63)$$

Proposition 6.1 (Extreme routing shares). *Assume $\bar{\theta}(0) = 0$ and $\bar{h}(0) \leq c$. If $\varrho = 1$, every equilibrium satisfies $\theta^* = 0$. If $\varrho = 0$, every equilibrium satisfies $h^* = 0$.*

Let

$$\underline{v} = \inf_{s \in [-Q_{\max}, Q_{\max}]} \bar{\lambda}(s), \quad \bar{v} = \sup_{s \in [-Q_{\max}, Q_{\max}]} \bar{\lambda}(s). \quad (64)$$

Theorem 6.2 (Primitive one-field coexistence and extinction). *Assume $\bar{h}$ and $\bar{\theta}$ are nondecreasing and $\varrho \in (0, 1)$.*

1. *If $\bar{\theta}((1 - \varrho)\underline{v}) > 0$ and*

$$c < \frac{\bar{h}(\varrho \underline{v})}{1 + \xi A_{\max}}, \quad (65)$$

then every equilibrium is mixed: $h^* > 0$ and $\theta^* > 0$.

2. *If*

$$c \geq \sup_{z \in [0, \bar{v}]} \bar{h}(\varrho z), \quad (66)$$

then every equilibrium satisfies $h^* = 0$.

The first part uses the lower viewer floor and the worst allowed AI mass, so it is conservative. The second part uses an upper human-reference bound and is also primitive.

Corollary 6.3 (Routing-share regions). *Assume the hypotheses of Theorem 6.2, $0 < \underline{v} \leq \bar{v}$, $\bar{h}$ is continuous and strictly increasing, and $\bar{\theta}(e) > 0$ for $e > e_A^0$, where $e_A^0 = \sup\{e \geq 0 : \bar{\theta}(e) = 0\}$. Let $\bar{h}^{\leftarrow}(y) = \inf\{e \geq 0 : \bar{h}(e) \geq y\}$, with the convention that the infimum is $+\infty$ when the set is empty. A primitive mixed-region certificate is*

$$\varrho > \frac{\bar{h}^{\leftarrow}(c(1 + \xi A_{\max}))}{\underline{v}} \quad \text{and} \quad \varrho < 1 - \frac{e_A^0}{\underline{v}}. \quad (67)$$

If $c \geq \bar{h}(0)$, a primitive human-extinction certificate is

$$\varrho \leq \min \left\{ 1, \frac{\bar{h}^{\leftarrow}(c)}{\bar{v}} \right\}. \quad (68)$$

Proposition 6.4 (More AI is not always better at fixed human mass). *Assume Assumption 3.2, $\varrho \in (0, 1)$, and $\tilde{f}^A$ is C^1 with nonnegative derivative. Fix $v > 0$ and $h > 0$, and define $\Theta_h(\theta) = vm(r(h, \theta))q(h, \theta)$. Then*

$$\Theta'_h(\theta) = vm(r(h, \theta)) \left[(1 - \varrho)(\tilde{f}^A)'(\theta) + 2\kappa(r(h, \theta) - \omega) \frac{h + \epsilon_r}{(h + \theta + 2\epsilon_r)^2} q(h, \theta) \right]. \quad (69)$$

If $\theta_\omega = (h + \epsilon_r)/\omega - h - 2\epsilon_r > 0$ and $q(h, \theta) > 0$ on $[0, \theta_\omega]$, then Θ_h is increasing on $[0, \theta_\omega]$. If, in addition, there are $\theta_+ \in [\theta_\omega, \bar{\theta})$ and $\bar{\theta} > \theta_\omega$ such that $\Theta'_h(\theta_+) > 0$ while the negative mixture term dominates the intrinsic AI-quality derivative at $\bar{\theta}$ in (69), then Θ_h has an interior local maximizer in $(\theta_+, \bar{\theta})$.

7 Experiments

7.1 Simulation protocol and diagnostics

The experiments are mechanism checks rather than empirical estimates. They use the same update equations as the theory with $K = L = 4$ viewer groups and creator fields. All random primitives are generated once from a fixed seed and then held fixed within a sweep unless the sweep explicitly changes a parameter. The human and intrinsic AI quality functions are the saturating maps

$$f_{k\ell}^H(h_\ell) = a_{k\ell}^H \{1 - \exp(-u_{k\ell}^H h_\ell)\}, \quad \tilde{f}_{k\ell}^A(\theta_\ell) = a_{k\ell}^A \{1 - \exp(-u_{k\ell}^A \theta_\ell)\}. \quad (70)$$

The reference-population maps are logistic caps,

$$\bar{\lambda}_k(s) = C_k^V \sigma(\beta_k^V (s - \tau_k^V)), \quad \bar{h}_\ell(e) = C_\ell^H \sigma(\beta_\ell^H (e - \tau_\ell^H)), \quad \bar{\theta}_\ell(e) = C_\ell^A \sigma(\beta_\ell^A (e - \tau_\ell^A)), \quad (71)$$

where $\sigma(z) = 1/(1 + e^{-z})$. For a trajectory $x_0, x_1, \dots, x_T$, each reported limiting statistic is the average over a terminal window. The numerical tail span is

$$\text{span}_{\text{tail}} = \max_{t \in \mathcal{T}} \left\| x_t - \frac{1}{|\mathcal{T}|} \sum_{u \in \mathcal{T}} x_u \right\|_2, \quad (72)$$

where $\mathcal{T}$ is the terminal window. A run is declared converged when the terminal one-step displacement and the mean period-one tail distance are below tolerance; a period- p run is declared when the mean lag- p tail distance is below tolerance while the lag-one distance is not. These labels are diagnostics only; the theory itself is stated in terms of fixed points and spectral conditions.

Unless otherwise specified, the baseline uses common relaxation $\eta_{\text{common}} = 0.7$, block relaxations $(\eta_V, \eta_H, \eta_A) = (0.385, 0.525, 0.665)$, mean human cost 0.269, mean crowd-out coefficient 0.228, mean target share 0.507, and mean human routing share 0.508. The medium initial condition is $v_0 = 0.401$, $h_0 = 0.701$, and $\theta_0 = 0.151$; the strike-state diagnostic instead sets one human field to activity level 1.0 and the remaining human fields to zero. The figures and tables in this section are exported from the simulation output included with the submission package. The executable notebook is CPU-only, uses the standard Python scientific stack, and does not train any model or require a GPU; each trajectory stores only the $K + 2L = 12$ state coordinates and scalar diagnostics, so the memory footprint is small and the full exported output bundle is under 5 MB. We report the ecology score $\mathcal{E}_{\epsilon_E}$, not the raw Shannon diversity component, because only the activity-adjusted quantity enters AI quality in (9) and because raw diversity can be visually misleading near human extinction.

The experiments cover the fixed-policy convergence result, the dynamic role of relaxation, the cost-induced support transition, the target-share derivative, and the ecology channel. The suite does not include an isolated one-field AI-mass sweep or an isolated single-field crowd-out sweep, so Proposition 6.4 and Theorem 5.3 remain analytic diagnostics in this paper.

Table 1: Main theory-to-simulation correspondences.

Theory result	Output	Observed pattern
Fixed-policy convergence	Figure 1, Table 2	Population paths converge and the residual span is 2.16×10^{-15} .
Relaxation and local stability	Figure 2, Table 4	Stable η values share the same limit; high η generates period-two and period-four cycles.
Cost support thresholds	Figure 3, Table 5	Multiplier 2.0 creates field-level dropout; multiplier 3.0 gives all-field dropout.
Target matching	Figure 4	Welfare peaks at $\omega = 0.40$, close to realized human share 0.391.
Ecology smoothing	Figure 5, Table 3	Balanced routing preserves more ecology; the ecology score changes nonmonotonically with ϵ_E near an almost extinct field.

7.2 Baseline convergence and policy comparison

Figure 1 shows the baseline fixed-policy trajectory. The state converges rapidly; welfare, satisfaction, field-level human shares, and the best-response gaps stabilize in a way consistent with the fixed-point equations. Table 2 reports the limiting summary. The platform sustains positive viewer, human, and AI populations, and its ecology score is 1.269. The realized mean human share is 0.391, below the

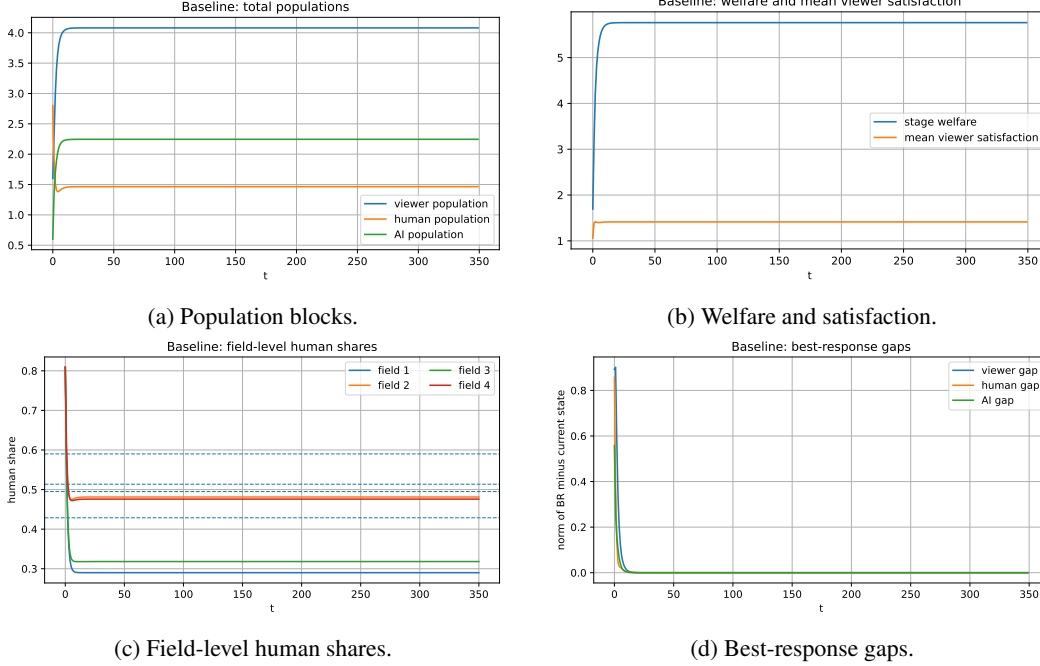


Figure 1: Baseline fixed-policy dynamics. The state converges to a steady state and the best-response residuals vanish.

Table 2: Baseline limiting statistics.

Status	Viewers	Human	AI	Welfare	Satisf.	Mix	Human share	Ecology
Converged	4.080	1.465	2.245	5.763	1.414	0.952	0.391	1.269

average target share 0.507. This gap is important for the target-share experiment below: when the target is set too high relative to the sustainable human share, the mixture factor penalizes welfare.

Table 3 compares three fixed policies. The hot policy maximizes stage welfare and viewer satisfaction in this synthetic instance, but it has lower ecology than the balanced-preference policy. The anti-hot policy preserves a similar ecology score but loses substantial welfare. The comparison illustrates why stage welfare, human support, and ecology should be diagnosed jointly.

7.3 Relaxation, instability, and cycles

The relaxation sweep directly illustrates Propositions 5.6 and Theorems 5.7–5.8. For $\eta \leq 0.90$, the welfare curve is essentially flat after transients: the fixed point is the same, while the path speed changes. At $\eta = 0.95$ and $\eta = 1.00$, the system no longer converges to the fixed point and instead exhibits a period-two tail. In the full η -slope grid, nonconvergence concentrates where both the relaxation and the reference-map slopes are large, which is the regime where negative local modes can be pushed past the Schur boundary.

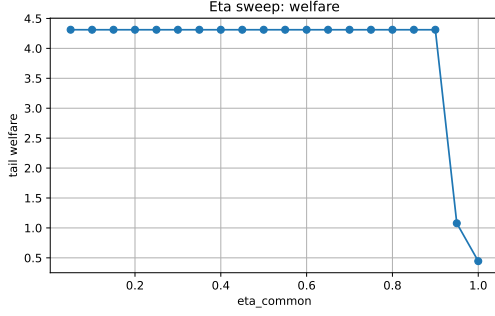
The high-slope diagnostic also produces a period-four orbit at $\eta = 1.00$ and slope scale 4.00. Across its four phases, aggregate human mass alternates between about 1.79 and zero, while viewer mass alternates between very low and high states. This is consistent with a delayed overreaction: when viewers surge, AI exposure expands and human support collapses; the next phase reverses the imbalance.

7.4 Human costs and support transitions

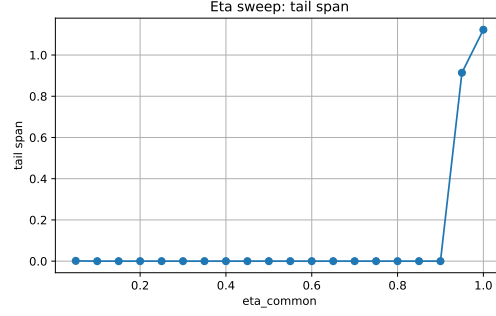
Figure 3 varies the human-cost multiplier. Welfare and human population decline monotonically in the sweep, while AI population declines more slowly. The field-level statistics in Table 5 show a clear

Table 3: Fixed-policy comparison.

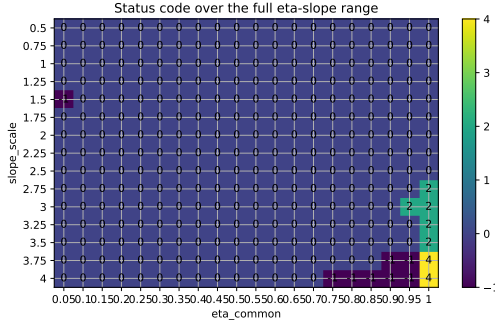
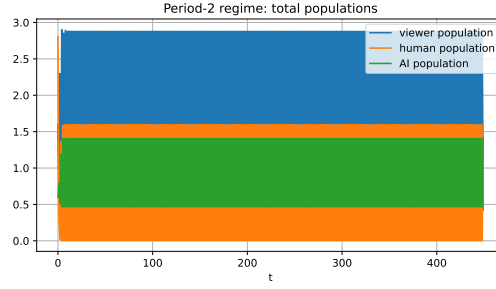
Policy	Viewers	Human	AI	Welfare	Satisf.	Ecology
Anti-hot	2.392	1.142	1.874	1.430	0.587	1.261
Balanced-pref	4.080	1.465	2.245	5.763	1.414	1.269
Hot	4.475	1.508	2.346	7.329	1.642	1.256



(a) Welfare versus relaxation.



(b) Tail span versus relaxation.

(c) Status over η and slope.

(d) Representative period-two path.

Figure 2: Relaxation controls dynamic stability. The fixed point is stable for moderate η , while high relaxation and steep reference maps produce cycles.

support transition. At multiplier 2.0, the minimum field-level human population is already zero while the maximum remains positive, so the system has entered a partial-dropout regime. At multiplier 3.0, all field-level human populations are numerically zero. The ecology score collapses with human activity, confirming that (9) measures an active ecosystem rather than merely a normalized distribution over fields.

7.5 Target human share

Figure 4 varies the common target share ω . The realized mean human share is stable near 0.39 over a broad range of targets. Consequently, the target-realization gap is smallest near $\omega = 0.40$, and both the mixture factor and welfare peak there. When ω is increased beyond what the equilibrium can support, the mixture factor declines sharply and pulls down welfare. This is the direct comparative-static mechanism in Proposition 5.5: the direct welfare derivative changes sign when the target crosses the realized share.

7.6 Ecology smoothing and near-extinction diagnostics

The ecology score in (9) combines field balance with total human activity. The strike-state diagnostic in Figure 5 varies ϵ_E near an initialization in which one human field is active and the other fields begin at zero. The status heatmap shows that the smoothing scale does not by itself cure aggressive relaxation: at $\eta_{\text{common}} = 1.00$ every sampled ϵ_E value remains period-two. Along the one-dimensional ϵ_E sweep, the tail ecology score is nonmonotone: it rises as smoothing regularizes the nearly extinct

Table 4: Aggregate period-two cycle at $\eta = 1.00$.

Block	Cycle point A	Cycle point B
Viewer	0.424	2.880
Human	1.597	0.000
AI	1.408	0.465

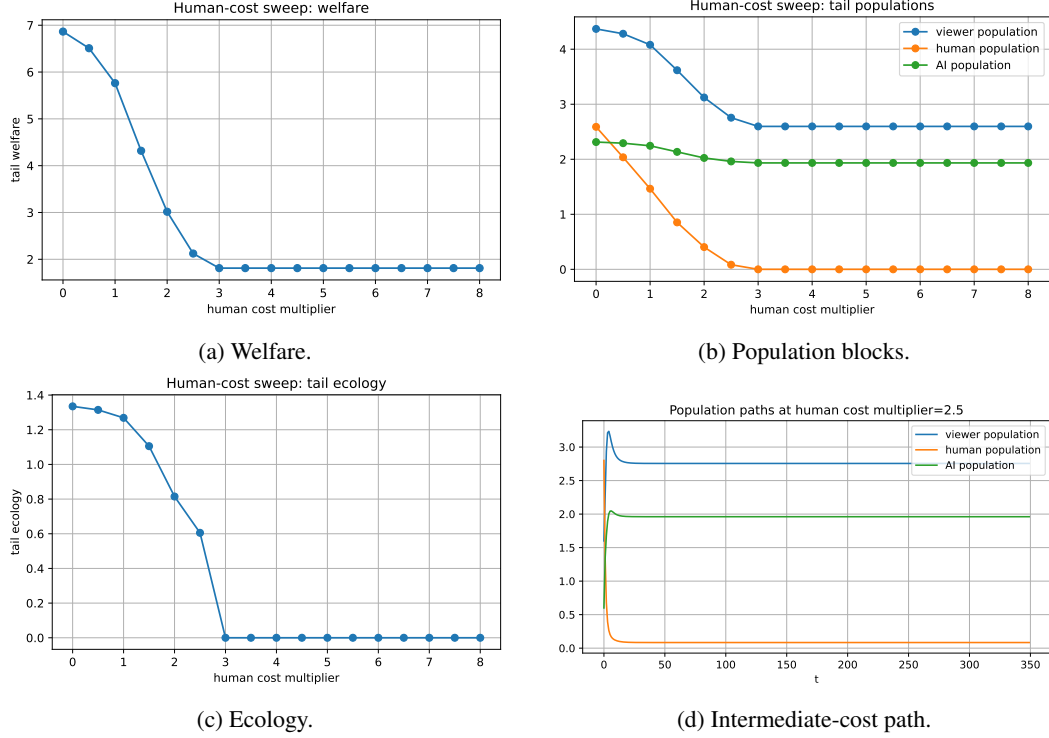


Figure 3: Human-cost sweep. Higher human cost first creates field-level dropout and then all-field human extinction.

fields, peaks around $\epsilon_E = 0.086$ with ecology about 0.391, and then falls as excessive smoothing dilutes the activity signal. The corresponding welfare curve changes much less smoothly because it also depends on population scale and quality. These diagnostics motivate reporting the ecology score itself: it is smooth enough for the theory, but it does not treat a nearly inactive human base as healthy.

Table 5: Selected cost-sweep support outcomes.

Cost mult.	Viewers	Human	AI	Min field h	Max field h
0.0	4.369	2.589	2.313	0.521	0.770
1.0	4.080	1.465	2.245	0.232	0.495
2.0	3.123	0.404	2.024	0.000	0.217
2.5	2.756	0.084	1.961	0.000	0.067
3.0	2.597	0.000	1.934	0.000	0.000
7.0	2.597	0.000	1.934	0.000	0.000

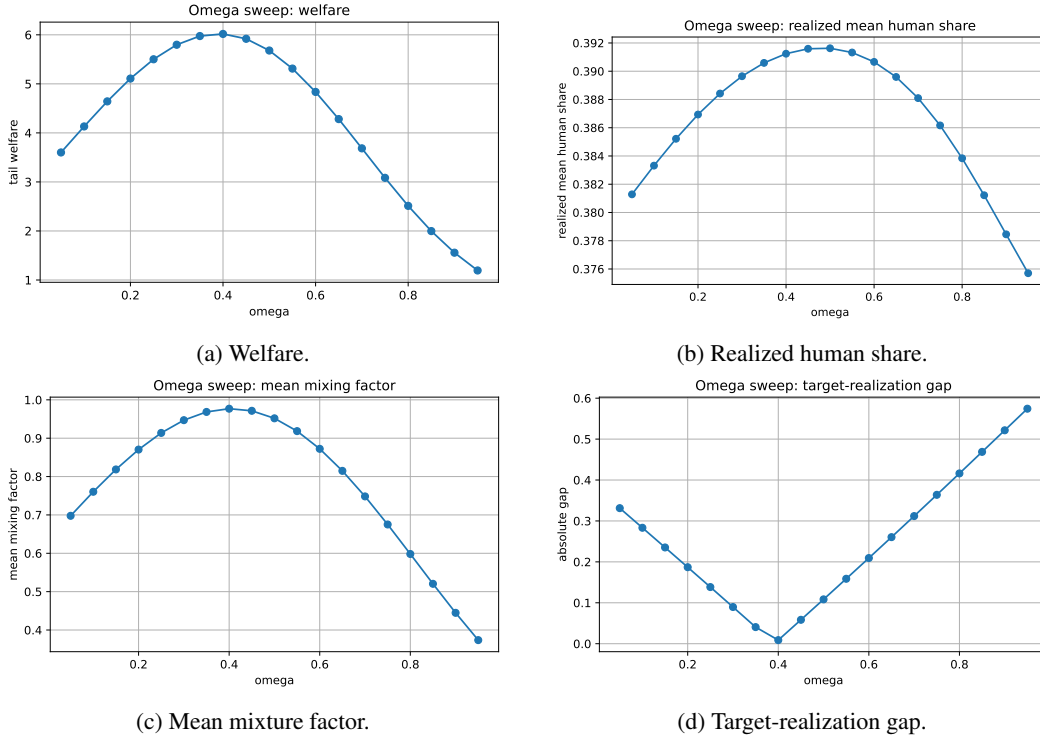


Figure 4: Common target-share sweep. Welfare peaks where the target is close to the realized human share.

Table 6: Selected target-share outcomes.

ω	Viewers	Human	AI	Welfare	Mean mix
0.20	3.886	1.406	2.217	5.108	0.870
0.35	4.142	1.467	2.264	5.975	0.968
0.40	4.153	1.474	2.264	6.018	0.977
0.55	3.942	1.447	2.210	5.312	0.918
0.80	2.912	1.258	1.981	2.512	0.598
0.95	2.236	1.124	1.840	1.195	0.374

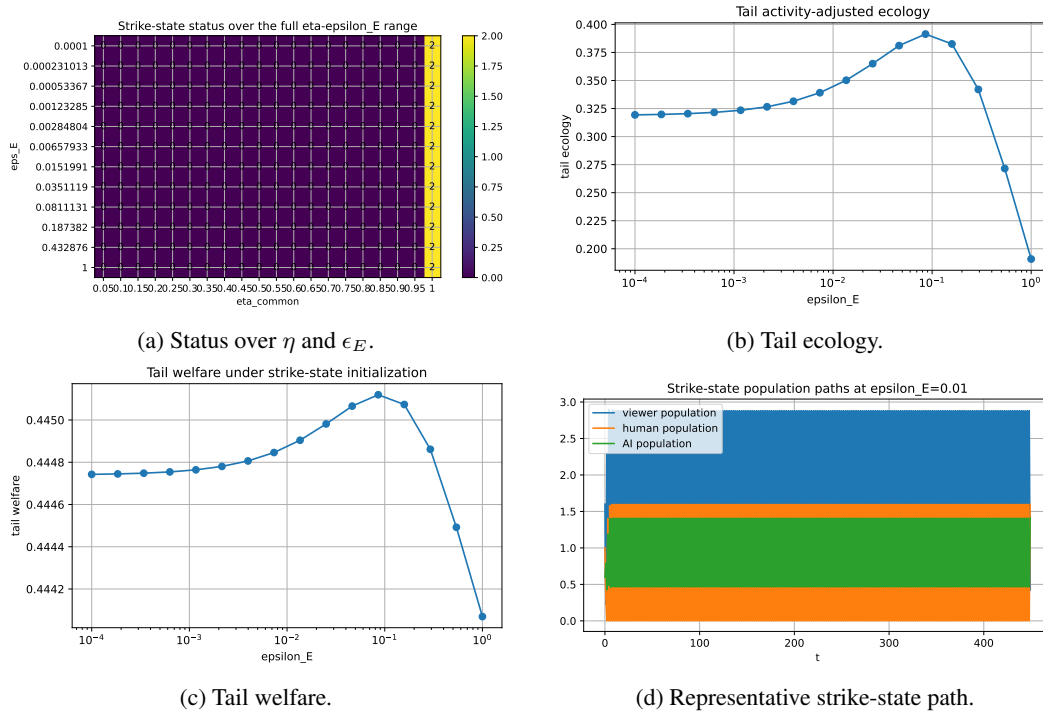


Figure 5: Ecology smoothing diagnostics. The reported ecology score is activity-adjusted and remains tied to the economically active human creator base.

8 Limitations and discussion

There are some limitations in this study. It focuses on fixed policies rather than solving a dynamic policy-optimization problem. The sufficient support certificates are conservative because they must hold uniformly for all equilibria in the compact domain. The experiments are synthetic and should be read as mechanism checks rather than empirical estimates. The deterministic mean-field dynamics abstract away from stochastic arrivals, creator heterogeneity within a field, strategic content adaptation, and platform learning. These abstractions make the fixed-point theory transparent, but they also leave several open directions: policy design under stability constraints, empirical calibration of crowd-out, stochastic versions of the dropout thresholds, and welfare criteria that explicitly price human ecology rather than allowing it to enter only through satisfaction.

The societal implication is two-sided. On the positive side, explicit participation dynamics can help platforms audit whether AI supply policies preserve human creator activity and ecological balance. On the negative side, the same model class could be used to tune recommendation or automation policies that increase short-run engagement while accelerating human creator crowd-out. The results should therefore be interpreted as a diagnostic framework for stability and participation risk, not as a recipe for suppressing human creators.

9 Conclusion

This paper develops a participation-dynamic model for two-sided platforms with human creators and AI shadow creators. The theory separates equilibrium support, global convergence, and relaxation-driven instability. Human dropout is controlled by primitive exposure, cost, and crowd-out thresholds. The common relaxation parameter changes convergence and local stability, not the equilibrium equations. The simulations show that AI supply can coexist with humans in stable regimes, but can also accelerate human dropout or produce cycles when the dynamic response is too aggressive. The main lesson is that recommendation policies for AI-rich platforms should be evaluated not only by current viewer satisfaction, but also by the participation states they induce on both sides of the market.

References

- [1] Mark Armstrong. Competition in two-sided markets. *RAND Journal of Economics*, 37(3): 668–691, 2006.
- [2] Omer Ben-Porat and Moshe Tennenholtz. A game-theoretic approach to recommendation systems with strategic content providers. In *Advances in Neural Information Processing Systems*, 2018.
- [3] Xiaowu Dai, Wenlu Xu, Yuan Qi, and Michael I. Jordan. Incentive-aware recommender systems in two-sided markets. *arXiv preprint arXiv:2211.15381*, 2022.
- [4] M. O. Hill. Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2): 427–432, 1973.
- [5] Jiri Hron, Karl Krauth, Michael I. Jordan, Niki Kilbertus, and Sarah Dean. Modeling content creator incentives on algorithm-curated platforms. In *International Conference on Learning Representations*, 2023.
- [6] Yiqing Hua, Shuo Niu, Jie Cai, Lydia B. Chilton, Hendrik Heuer, and Donghee Yvette Wohn. Generative ai in user-generated content. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 2024.
- [7] Daniel Huttenlocher, Hannah Li, Liang Lyu, Asuman Ozdaglar, and James Siderius. Matching of users and creators in two-sided markets with departures. *arXiv preprint arXiv:2401.00313*, 2024.
- [8] Ray Jiang, Silvia Chiappa, Tor Lattimore, Andras Gyorgy, and Pushmeet Kohli. Degenerate feedback loops in recommender systems. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 383–390, 2019.

- [9] Lou Jost. Entropy and diversity. *Oikos*, 113(2):363–375, 2006.
- [10] Sueyoul Kim, Ginger Zhe Jin, and Eungik Lee. Does generative ai crowd out human creators? evidence from pixiv. Technical Report 34733, National Bureau of Economic Research, 2026.
- [11] Haruka Kiyohara, Fan Yao, and Sarah Dean. Policy design for two-sided platforms with participation dynamics. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025.
- [12] Gourab K. Patro, Arpita Biswas, Niloy Ganguly, Krishna P. Gummadi, and Abhijnan Chakraborty. FairRec: Two-sided fairness for personalized recommendations in two-sided platforms. In *Proceedings of The Web Conference 2020*, pages 1194–1204, 2020.
- [13] E. C. Pielou. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, 13:131–144, 1966.
- [14] Jean-Charles Rochet and Jean Tirole. Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029, 2003.
- [15] Marc Rysman. The economics of two-sided markets. *Journal of Economic Perspectives*, 23(3): 125–143, 2009.
- [16] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.
- [17] Ashudeep Singh and Thorsten Joachims. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2219–2228, 2018.
- [18] Fan Yao, Chuanhao Li, Denis Nekipelov, Hongning Wang, and Haifeng Xu. How bad is top- k recommendation under competing content creators? In *Proceedings of the 40th International Conference on Machine Learning*, 2023.
- [19] Yukun Zhang. The influence of generative ai on content platforms: Supply, demand, and welfare impacts in two-sided markets. *arXiv preprint arXiv:2410.13101*, 2024.

A Proofs

This appendix gives the complete mathematical verification of all formal claims in the main text. Throughout, π is fixed unless a proof explicitly varies it, and all norms are Euclidean norms except for the weighted block-max norm introduced in the proof of Theorem 4.6. Every theorem, proposition, lemma, and corollary stated in the main text has a corresponding proof below with the same label.

A.1 Preliminary bounds used repeatedly

We first record two elementary facts that are used in several proofs. Assumption 3.1 implies that every best-response coordinate is bounded by the corresponding side of the compact box $\mathcal{X}$. Indeed,

$$0 \leq \text{BR}_k^V(x; \pi) = \bar{\lambda}_k(s_k(h, \theta; \pi)) \leq V_{\max}, \quad (73)$$

$$0 \leq \text{BR}_\ell^H(x; \pi) = \left[\frac{\bar{h}_\ell(e_\ell^H(v; \pi))}{1 + \xi_\ell \theta_\ell} - c_\ell \right]_+ \leq \frac{\bar{h}_\ell(e_\ell^H(v; \pi))}{1 + \xi_\ell \theta_\ell} \leq H_{\max}, \quad (74)$$

and

$$0 \leq \text{BR}_\ell^A(x; \pi) = \bar{\theta}_\ell(e_\ell^A(v; \pi)) \leq A_{\max}. \quad (75)$$

Thus $\text{BR}(\mathcal{X}; \pi) \subseteq \mathcal{X}$.

Second, every exposure coordinate generated by a state in $\mathcal{X}$ lies in $[0, E_{\max}]$. Since $0 \leq \varrho_{k\ell} \leq 1$, $0 \leq \pi_{k\ell} \leq 1$, and $0 \leq v_k \leq V_{\max}$,

$$0 \leq e_\ell^H(v; \pi) \leq \sum_{k=1}^K v_k \leq KV_{\max} = E_{\max}, \quad 0 \leq e_\ell^A(v; \pi) \leq \sum_{k=1}^K v_k \leq E_{\max}. \quad (76)$$

The definition of $Q_{\max}$ in (20), the facts $0 < m_\ell \leq 1$, $0 \leq a_{\epsilon_E} \leq 1$, and $0 \leq \mathcal{D}_{\epsilon_E} \leq \log L$, imply

$$|Q_{k\ell}(h, \theta)| \leq Q_{\max}, \quad |s_k(h, \theta; \pi)| \leq Q_{\max} \quad (77)$$

for every (h, θ) in the projection of $\mathcal{X}$ and every row-stochastic policy π .

A.2 Proofs for Sections 3 and 4

Proof of Proposition 3.5. Fix $(\pi, x) \in \Pi \times \mathcal{X}$ and write $x = (v, h, \theta)$. For each field ℓ ,

$$h_\ell + \theta_\ell + 2\epsilon_r \geq 2\epsilon_r > 0, \quad (78)$$

so the smoothed share

$$r_\ell(h, \theta) = \frac{h_\ell + \epsilon_r}{h_\ell + \theta_\ell + 2\epsilon_r} \quad (79)$$

is well defined and lies in $(0, 1)$. Hence $m_\ell(r_\ell)$ is finite and strictly positive. Similarly,

$$\sum_{i=1}^L (h_i + \epsilon_E) \geq L\epsilon_E > 0, \quad (80)$$

so the smoothed probabilities $p_j(h)$ are well defined and strictly positive. Therefore $p_j \log p_j$, $\mathcal{D}_{\epsilon_E}(h)$, $a_{\epsilon_E}(h)$, and $\mathcal{E}_{\epsilon_E}(h)$ are all well defined. The data map $\log(1 + h_\ell)$ is well defined for $h_\ell \geq 0$, and all primitive quality maps are continuous on their domains. Thus each $Q_{k\ell}(h, \theta)$, each satisfaction coordinate $s_k(h, \theta; \pi)$, and finally

$$W(\pi; x) = \sum_{k=1}^K v_k s_k(h, \theta; \pi) \quad (81)$$

are well defined.

Now impose Assumption 3.2. For every $h \in \mathbb{R}_+^L$, $d_\ell(h_\ell) = \log(1 + h_\ell) \geq 0$. Also $\mathcal{D}_{\epsilon_E}(h) \geq 0$ and $a_{\epsilon_E}(h) \geq 0$, so $\mathcal{E}_{\epsilon_E}(h) \geq 0$. Hence

$$f_{k\ell}^A(\theta_\ell; h_\ell, h) = \tilde{f}_{k\ell}^A(\theta_\ell) + \gamma_{k\ell}^{\text{data}} \log(1 + h_\ell) + \gamma^{\text{ent}} \mathcal{E}_{\epsilon_E}(h) \geq 0. \quad (82)$$

Together with $b_{k\ell} \geq 0$ and $f_{k\ell}^H \geq 0$, this gives $Q_{k\ell}(h, \theta) \geq 0$. Since $\pi_{k\ell} \geq 0$ and $m_\ell(r_\ell) > 0$, each summand in s_k is nonnegative, and $s_k \geq 0$. Finally $v_k \geq 0$ for $x \in \mathcal{X}$, so each term $v_k s_k$ in W is nonnegative. Therefore $W(\pi; x) \geq 0$. $\square$

Proof of Proposition 4.1. Fix π . In the induced game, each player controls one scalar coordinate and treats the other coordinates as fixed.

For viewer group k , the quantity $s_k(h, \theta; \pi)$ does not depend on the viewer's own coordinate v_k . Let

$$a_k = \bar{\lambda}_k(s_k(h, \theta; \pi)) \geq 0. \quad (83)$$

The viewer's own-coordinate problem is

$$\max_{z \geq 0} a_k z - \frac{z^2}{2}. \quad (84)$$

The objective is strictly concave, with derivative $a_k - z$, so its unconstrained maximizer is $z = a_k$. Since $a_k \geq 0$, this is also the constrained maximizer. Thus the unique viewer best response is BR_k^V .

For human field ℓ , $e_\ell^H(v; \pi)$ does not depend on h_ℓ . With

$$b_\ell = \frac{\bar{h}_\ell(e_\ell^H(v; \pi))}{1 + \xi_\ell \theta_\ell} - c_\ell, \quad (85)$$

the human field solves

$$\max_{z \geq 0} b_\ell z - \frac{z^2}{2}. \quad (86)$$

The unconstrained maximizer is $z = b_\ell$. Projecting onto the feasible half-line gives the unique constrained maximizer $[b_\ell]_+$, which is BR_ℓ^H .

For AI field ℓ , $e_\ell^A(v; \pi)$ does not depend on θ_ℓ . With

$$a_\ell^A = \bar{\theta}_\ell(e_\ell^A(v; \pi)) \geq 0, \quad (87)$$

the own-coordinate problem is

$$\max_{z \geq 0} a_\ell^A z - \frac{z^2}{2}. \quad (88)$$

The unique constrained maximizer is $z = a_\ell^A$, hence the AI best response is BR_ℓ^A .

A Nash equilibrium is therefore precisely a state at which every coordinate equals its unique best response, i.e. $x^* = \text{BR}(x^*; \pi)$.

For the dynamical statement, use the relaxed update

$$g(x, \pi) = x + H\{\text{BR}(x; \pi) - x\}. \quad (89)$$

The diagonal matrix H has strictly positive diagonal entries because $\eta_V, \eta_H, \eta_A \in (0, 1]$. Consequently,

$$g(x^*, \pi) = x^* \iff H\{\text{BR}(x^*; \pi) - x^*\} = 0 \iff \text{BR}(x^*; \pi) = x^*. \quad (90)$$

Thus Nash equilibria, best-response fixed points, and steady states coincide. $\square$

Proof of Theorem 4.2. The preliminary bounds show that $\text{BR}(\mathcal{X}; \pi) \subseteq \mathcal{X}$. If $x \in \mathcal{X}$, then

$$g(x, \pi) = (I - H)x + H \text{BR}(x; \pi) \quad (91)$$

is a coordinatewise convex combination of x and $\text{BR}(x; \pi)$, with weights in $[0, 1]$. Since $\mathcal{X}$ is a product of closed intervals and therefore convex, $g(x, \pi) \in \mathcal{X}$. This proves $g(\mathcal{X}, \pi) \subseteq \mathcal{X}$.

The map $g(\cdot, \pi)$ is continuous on $\mathcal{X}$. To see this, note that e^H and e^A are linear in v ; the denominators $h_\ell + \theta_\ell + 2\epsilon_r$ and $1 + \xi_\ell \theta_\ell$ are bounded away from zero; the maps p_j , $\mathcal{D}_{\epsilon_E}$, a_{ϵ_E} , $\mathcal{E}_{\epsilon_E}$, m_ℓ , and $[\cdot]_+$ are continuous; and the primitive maps are continuous by Assumption 3.1. Hence $\text{BR}(\cdot; \pi)$ and $g(\cdot, \pi)$ are continuous.

The set $\mathcal{X}$ is nonempty, compact, and convex in the finite-dimensional Euclidean space $\mathbb{R}^{K+2L}$. Brouwer's fixed point theorem applied to the continuous self-map $g(\cdot, \pi) : \mathcal{X} \rightarrow \mathcal{X}$ gives at least one fixed point in $\mathcal{X}$.

Forward invariance follows by induction. If $x_t \in \mathcal{X}$, then $x_{t+1} = g(x_t, \pi) \in \mathcal{X}$. Since $x_0 \in \mathcal{X}$, every subsequent state remains in $\mathcal{X}$. $\square$

Proof of Proposition 4.3. Fix $x = (v, h, \theta) \in \mathcal{X}$. Let

$$z_\ell = \frac{\bar{h}_\ell(e_\ell^H(v; \pi))}{1 + \xi_\ell \theta_\ell} - c_\ell. \quad (92)$$

The human best response is $\text{BR}_\ell^H(x; \pi) = [z_\ell]_+$. Since $[z]_+ = 0$ if and only if $z \leq 0$,

$$\text{BR}_\ell^H(x; \pi) = 0 \iff \frac{\bar{h}_\ell(e_\ell^H(v; \pi))}{1 + \xi_\ell \theta_\ell} \leq c_\ell. \quad (93)$$

The denominator is strictly positive. Multiplying both sides by $1 + \xi_\ell \theta_\ell$ gives

$$\bar{h}_\ell(e_\ell^H(v; \pi)) \leq c_\ell(1 + \xi_\ell \theta_\ell), \quad (94)$$

which proves the equivalence. The right side is weakly increasing in c_ℓ and in ξ_ℓ because $\theta_\ell \geq 0$, so increasing either parameter weakly enlarges the dropout region. $\square$

Proof of Lemma 4.4. We prove the three building-block estimates separately.

Step 1: smooth share and mixture factor. For a fixed field ℓ , set

$$r_\ell(h, \theta) = \frac{h_\ell + \epsilon_r}{h_\ell + \theta_\ell + 2\epsilon_r}. \quad (95)$$

On $[0, H_{\max}] \times [0, A_{\max}]$,

$$\frac{\partial r_\ell}{\partial h_\ell} = \frac{\theta_\ell + \epsilon_r}{(h_\ell + \theta_\ell + 2\epsilon_r)^2}, \quad (96)$$

$$\frac{\partial r_\ell}{\partial \theta_\ell} = -\frac{h_\ell + \epsilon_r}{(h_\ell + \theta_\ell + 2\epsilon_r)^2}. \quad (97)$$

Because $h_\ell + \theta_\ell + 2\epsilon_r \geq 2\epsilon_r$, the absolute values of these derivatives are bounded by

$$C_{r,H} = \frac{A_{\max} + \epsilon_r}{4\epsilon_r^2}, \quad C_{r,A} = \frac{H_{\max} + \epsilon_r}{4\epsilon_r^2}. \quad (98)$$

The mean-value theorem yields

$$|r_\ell(h, \theta) - r_\ell(h', \theta')| \leq C_{r,H} |h_\ell - h'_\ell| + C_{r,A} |\theta_\ell - \theta'_\ell|. \quad (99)$$

The function $m_\ell(r) = \exp\{-\kappa_\ell(r - \omega_\ell)^2\}$ is C^1 on $[0, 1]$. Let

$$L_m = \max_{\ell \leq L} \sup_{r \in [0, 1]} 2\kappa_\ell |r - \omega_\ell| \exp\{-\kappa_\ell(r - \omega_\ell)^2\} < \infty. \quad (100)$$

Then

$$|m_\ell(r_\ell) - m_\ell(r'_\ell)| \leq L_m |r_\ell - r'_\ell|. \quad (101)$$

Step 2: quality aggregate. The map $d(z) = \log(1 + z)$ is 1-Lipschitz on $\mathbb{R}_+$ because $d'(z) = 1/(1 + z) \leq 1$. The map $\mathcal{E}_{\epsilon_E}$ is C^1 on the compact cube $[0, H_{\max}]^L$: the denominator in $p_j(h)$ is at least $L\epsilon_E$, so no singularity occurs. Hence

$$L_E := \sup_{h \in [0, H_{\max}]^L} \|D\mathcal{E}_{\epsilon_E}(h)\| < \infty \quad (102)$$

and

$$|\mathcal{E}_{\epsilon_E}(h) - \mathcal{E}_{\epsilon_E}(h')| \leq L_E \|h - h'\|. \quad (103)$$

Using Assumption 3.3, for every (k, ℓ) there are finite constants $C_{k\ell}^H$ and $C_{k\ell}^A$ such that

$$|Q_{k\ell}(h, \theta) - Q_{k\ell}(h', \theta')| \leq C_{k\ell}^H \|h - h'\| + C_{k\ell}^A \|\theta - \theta'\|. \quad (104)$$

For example, one may take

$$C_{k\ell}^H = \varrho_{k\ell} L_{fH} + (1 - \varrho_{k\ell}) \{\gamma_{k\ell}^{\text{data}} + \gamma^{\text{ent}} L_E\}, \quad (105)$$

$$C_{k\ell}^A = (1 - \varrho_{k\ell}) L_{fA}. \quad (106)$$

Let $C_Q^H = \max_{k\ell} C_{k\ell}^H$ and $C_Q^A = \max_{k\ell} C_{k\ell}^A$.

Step 3: satisfaction. For one row k ,

$$\begin{aligned}
& |s_k(h, \theta; \pi) - s_k(h', \theta'; \pi)| \\
& \leq \sum_{\ell=1}^L \pi_{k\ell} |m_\ell(r_\ell) Q_{k\ell}(h, \theta) - m_\ell(r'_\ell) Q_{k\ell}(h', \theta')| \\
& \leq \sum_{\ell=1}^L \pi_{k\ell} \{ |m_\ell(r_\ell) - m_\ell(r'_\ell)| |Q_{k\ell}(h, \theta)| + |Q_{k\ell}(h, \theta) - Q_{k\ell}(h', \theta')| \} \\
& \leq \sum_{\ell=1}^L \pi_{k\ell} \{ Q_{\max} L_m (C_{r,H} |h_\ell - h'_\ell| + C_{r,A} |\theta_\ell - \theta'_\ell|) + C_Q^H \|h - h'\| + C_Q^A \|\theta - \theta'\| \}. \quad (107)
\end{aligned}$$

Using $\sum_{\ell} \pi_{k\ell} = 1$, $|h_\ell - h'_\ell| \leq \|h - h'\|$, and $|\theta_\ell - \theta'_\ell| \leq \|\theta - \theta'\|$, we obtain row-wise constants $\tilde{L}_{s,H}$ and $\tilde{L}_{s,A}$ such that

$$|s_k(h, \theta; \pi) - s_k(h', \theta'; \pi)| \leq \tilde{L}_{s,H} \|h - h'\| + \tilde{L}_{s,A} \|\theta - \theta'\|. \quad (108)$$

Multiplying by $\sqrt{K}$ gives finite constants $L_{s,H}$ and $L_{s,A}$ satisfying (27).

Step 4: exposure and effective human reference. The exposure maps are linear:

$$e^H(v; \pi) = A^H(\pi)v, \quad e^A(v; \pi) = A^A(\pi)v. \quad (109)$$

Therefore

$$\begin{aligned}
\|e^H(v; \pi) - e^H(v'; \pi)\| & \leq \|A^H(\pi)\|_2 \|v - v'\|, \\
\|e^A(v; \pi) - e^A(v'; \pi)\| & \leq \|A^A(\pi)\|_2 \|v - v'\|. \quad (110)
\end{aligned}$$

Finally, for $e, e' \in [0, E_{\max}]$ and $\theta, \theta' \in [0, A_{\max}]$,

$$\begin{aligned}
\left| \frac{\bar{h}_\ell(e)}{1 + \xi_\ell \theta} - \frac{\bar{h}_\ell(e')}{1 + \xi_\ell \theta'} \right| & \leq \frac{|\bar{h}_\ell(e) - \bar{h}_\ell(e')|}{1 + \xi_\ell \theta} + \bar{h}_\ell(e') \left| \frac{1}{1 + \xi_\ell \theta} - \frac{1}{1 + \xi_\ell \theta'} \right| \\
& \leq L_h |e - e'| + H_{\max} \frac{\xi_\ell |\theta - \theta'|}{(1 + \xi_\ell \theta)(1 + \xi_\ell \theta')} \\
& \leq L_h |e - e'| + \Xi_{\max} H_{\max} |\theta - \theta'|. \quad (111)
\end{aligned}$$

This proves all claimed bounds. $\square$

Proof of Lemma 4.5. Write $x = (v, h, \theta)$ and $x' = (v', h', \theta')$. The viewer block of the relaxed update is

$$g_V(x) = (1 - \eta_V)v + \eta_V \text{BR}^V(x; \pi). \quad (112)$$

Hence

$$\begin{aligned}
\|g_V(x) - g_V(x')\| & \leq (1 - \eta_V) \|v - v'\| + \eta_V \|\text{BR}^V(x; \pi) - \text{BR}^V(x'; \pi)\| \\
& \leq (1 - \eta_V) \|v - v'\| + \eta_V L_\lambda \|s(h, \theta; \pi) - s(h', \theta'; \pi)\| \\
& \leq (1 - \eta_V) \|v - v'\| + \eta_V L_\lambda L_{s,H} \|h - h'\| + \eta_V L_\lambda L_{s,A} \|\theta - \theta'\|. \quad (113)
\end{aligned}$$

This is the first row of $M(\pi)\Delta(x, x')$.

For the human block,

$$g_H(x) = (1 - \eta_H)h + \eta_H \text{BR}^H(x; \pi). \quad (114)$$

The projection $z \mapsto [z]_+$ is nonexpansive on $\mathbb{R}$, so, coordinate by coordinate and then in Euclidean norm,

$$\begin{aligned}
\|\text{BR}^H(x; \pi) - \text{BR}^H(x'; \pi)\| & \leq L_h \|e^H(v; \pi) - e^H(v'; \pi)\| + \Xi_{\max} H_{\max} \|\theta - \theta'\| \\
& \leq L_h \|A^H(\pi)\|_2 \|v - v'\| + \Xi_{\max} H_{\max} \|\theta - \theta'\|. \quad (115)
\end{aligned}$$

Therefore

$$\|g_H(x) - g_H(x')\| \leq \eta_H L_h \|A^H(\pi)\|_2 \|v - v'\| + (1 - \eta_H) \|h - h'\| + \eta_H \Xi_{\max} H_{\max} \|\theta - \theta'\|, \quad (116)$$

which is the second row.

For the AI block,

$$\|\text{BR}^A(x; \pi) - \text{BR}^A(x'; \pi)\| \leq L_\theta \|e^A(v; \pi) - e^A(v'; \pi)\| \leq L_\theta \|A^A(\pi)\|_2 \|v - v'\|. \quad (117)$$

Thus

$$\|g_A(x) - g_A(x')\| \leq \eta_A L_\theta \|A^A(\pi)\|_2 \|v - v'\| + (1 - \eta_A) \|\theta - \theta'\|, \quad (118)$$

which is the third row. Combining the three block inequalities gives the componentwise comparison inequality. $\square$

Proof of Theorem 4.6. Theorem 4.2 already gives at least one fixed point in $\mathcal{X}$. We prove uniqueness and global convergence.

Let $M = M(\pi)$. Since $M \geq 0$ and $\rho(M) < 1$, the Neumann series converges and

$$w = (I_3 - M)^{-1} \mathbf{1} = \sum_{n=0}^{\infty} M^n \mathbf{1} \quad (119)$$

is a vector in $\mathbb{R}_{++}^3$. Moreover,

$$Mw = w - \mathbf{1} < w \quad (120)$$

componentwise. Define

$$\alpha = \max_{j=1,2,3} \frac{(Mw)_j}{w_j}. \quad (121)$$

Then $0 \leq \alpha < 1$ and $Mw \leq \alpha w$.

Define the weighted block-max norm

$$\|x - x'\|_w = \max_{j=1,2,3} \frac{\Delta_j(x, x')}{w_j}. \quad (122)$$

If $\|x - x'\|_w \leq c$, then $\Delta(x, x') \leq cw$ componentwise. Lemma 4.5 gives

$$\Delta(g(x, \pi), g(x', \pi)) \leq M \Delta(x, x') \leq cMw \leq c\alpha w. \quad (123)$$

Therefore

$$\|g(x, \pi) - g(x', \pi)\|_w \leq \alpha \|x - x'\|_w. \quad (124)$$

Thus $g(\cdot, \pi)$ is a contraction on the complete metric space $(\mathcal{X}, \|\cdot\|_w)$. Banach's fixed point theorem yields a unique fixed point $x^*(\pi)$ and, for every $x_0 \in \mathcal{X}$,

$$\|x_t - x^*(\pi)\|_w \leq \alpha^t \|x_0 - x^*(\pi)\|_w. \quad (125)$$

This is geometric convergence.

Finally, suppose a nontrivial periodic orbit of period $p \geq 2$ existed. Then g^p would fix every point on that orbit. But g^p is a contraction with factor at most $\alpha^p < 1$, so it has a unique fixed point. Since $x^*(\pi)$ is fixed by both g and g^p , the unique fixed point of g^p is $x^*(\pi)$. Hence no other point can lie on a periodic orbit. This proves the final claim. $\square$

A.3 Proofs for two-field results

Proof of Theorem 5.1. Let $x^* = (v^*, h^*, \theta^*)$ be any fixed-policy equilibrium. Since $x^* = \text{BR}(x^*; \pi)$, the viewer coordinates satisfy

$$v_k^* = \bar{\lambda}_k(s_k^*), \quad s_k^* = s_k(h^*, \theta^*; \pi). \quad (126)$$

The preliminary satisfaction bound gives $s_k^* \in [-Q_{\max}, Q_{\max}]$. By definition of $\underline{v}_k$,

$$v_k^* \geq \underline{v}_k, \quad k = 1, 2. \quad (127)$$

Therefore, for every field ℓ ,

$$e_\ell^{H,*} = \sum_{k=1}^2 \varrho_{k\ell} \pi_{k\ell} v_k^* \geq \sum_{k=1}^2 \varrho_{k\ell} \pi_{k\ell} v_k = \underline{e}_\ell^H(\pi), \quad (128)$$

because all coefficients are nonnegative. Also $0 \leq \theta_\ell^* \leq A_{\max}$ because $x^* \in \mathcal{X}$.

First consider a field $\ell \in S$. Since $\bar{h}_\ell$ is nondecreasing,

$$\bar{h}_\ell(e_\ell^{H,*}) \geq \bar{h}_\ell(\underline{e}_\ell^H(\pi)). \quad (129)$$

Since $\theta_\ell^* \leq A_{\max}$,

$$1 + \xi_\ell \theta_\ell^* \leq 1 + \xi_\ell A_{\max}. \quad (130)$$

Consequently,

$$\frac{\bar{h}_\ell(e_\ell^{H,*})}{1 + \xi_\ell \theta_\ell^*} \geq \frac{\bar{h}_\ell(\underline{e}_\ell^H(\pi))}{1 + \xi_\ell A_{\max}} = A_\ell(\pi). \quad (131)$$

If $c_\ell < A_\ell(\pi)$, then the pre-projection human best response is strictly positive:

$$\frac{\bar{h}_\ell(e_\ell^{H,*})}{1 + \xi_\ell \theta_\ell^*} - c_\ell > 0. \quad (132)$$

Thus $h_\ell^* = \text{BR}_\ell^H(x^*; \pi) > 0$.

Now consider a field $\ell \notin S$. For any equilibrium, $e_\ell^{H,*} \in [0, E_{\max}]$, so

$$\bar{h}_\ell(e_\ell^{H,*}) \leq U_\ell. \quad (133)$$

If $c_\ell \geq U_\ell$, then, because $\xi_\ell \theta_\ell^* \geq 0$,

$$\bar{h}_\ell(e_\ell^{H,*}) \leq U_\ell \leq c_\ell \leq c_\ell(1 + \xi_\ell \theta_\ell^*). \quad (134)$$

The dropout condition in Proposition 4.3 gives $h_\ell^* = 0$.

The four displayed regimes are obtained by substituting $S = \{1, 2\}$, $S = \{1\}$, $S = \{2\}$, and $S = \emptyset$ into the general statement. $\square$

Proof of Corollary 5.2. With $c_\ell(\alpha) = \alpha c_\ell^0$ and $c_\ell^0 > 0$, the active certificate $c_\ell(\alpha) < A_\ell(\pi)$ is equivalent to

$$\alpha < \frac{A_\ell(\pi)}{c_\ell^0} = a_\ell. \quad (135)$$

The inactive certificate $c_\ell(\alpha) \geq U_\ell$ is equivalent to

$$\alpha \geq \frac{U_\ell}{c_\ell^0} = u_\ell. \quad (136)$$

Applying Theorem 5.1 to $S = \{1, 2\}$ gives both fields active when $\alpha < a_1$ and $\alpha < a_2$, i.e. $\alpha < \min\{a_1, a_2\}$. Applying it to $S = \{1\}$ gives only field 1 active when $\alpha < a_1$ and $\alpha \geq u_2$. Applying it to $S = \{2\}$ gives only field 2 active when $\alpha < a_2$ and $\alpha \geq u_1$. Applying it to $S = \emptyset$ gives no human field active when $\alpha \geq u_1$ and $\alpha \geq u_2$, i.e. $\alpha \geq \max\{u_1, u_2\}$. If a one-field interval has left endpoint at least its right endpoint, no α can satisfy both inequalities, which proves the final statement. $\square$

Proof of Theorem 5.3. Fix the policy and all primitives except ξ_j .

For the active certificate, suppose $0 \leq \xi_j < \xi_j^{\text{on}}(\pi)$. If $\xi_j^{\text{on}}(\pi) \leq 0$, this condition is impossible and the statement is vacuous. Otherwise the inequality is equivalent to

$$c_j(1 + \xi_j A_{\max}) < B_j(\pi) = \bar{h}_j(\underline{e}_j^H(\pi)). \quad (137)$$

At any equilibrium, $e_j^{H,*} \geq \underline{e}_j^H(\pi)$ and $\theta_j^* \leq A_{\max}$. Monotonicity of $\bar{h}_j$ and (137) give

$$\frac{\bar{h}_j(e_j^{H,*})}{1 + \xi_j \theta_j^*} \geq \frac{\bar{h}_j(\underline{e}_j^H(\pi))}{1 + \xi_j A_{\max}} > c_j. \quad (138)$$

Therefore the human best-response pre-projection term is strictly positive, so $h_j^* > 0$ at every equilibrium.

For the inactive certificate, first suppose $\underline{\theta}_j(\pi) > 0$ and $\xi_j \geq \xi_j^{\text{off}}(\pi)$. By the definition of ξ_j^{off} , this implies

$$U_j \leq c_j(1 + \xi_j \underline{\theta}_j(\pi)). \quad (139)$$

At equilibrium,

$$\theta_j^* = \bar{\theta}_j(e_j^{A,*}). \quad (140)$$

As above, $v_k^* \geq v_k$, so $e_j^{A,*} \geq e_j^A(\pi)$. Since $\bar{\theta}_j$ is nondecreasing,

$$\theta_j^* \geq \bar{\theta}_j(e_j^A(\pi)) = \underline{\theta}_j(\pi). \quad (141)$$

Also $\bar{h}_j(e_j^{H,*}) \leq U_j$. Combining these inequalities with (139),

$$\bar{h}_j(e_j^{H,*}) \leq U_j \leq c_j(1 + \xi_j \underline{\theta}_j(\pi)) \leq c_j(1 + \xi_j \theta_j^*). \quad (142)$$

By Proposition 4.3, $h_j^* = 0$ at every equilibrium.

If $\underline{\theta}_j(\pi) = 0$ and $c_j \geq U_j$, the definition sets $\xi_j^{\text{off}}(\pi) = 0$. For any $\xi_j \geq 0$ and any equilibrium,

$$\bar{h}_j(e_j^{H,*}) \leq U_j \leq c_j \leq c_j(1 + \xi_j \theta_j^*), \quad (143)$$

so field j is inactive. If $\underline{\theta}_j(\pi) = 0$ and $c_j < U_j$, the definition sets $\xi_j^{\text{off}}(\pi) = +\infty$, so no finite ξ_j satisfies the off-threshold condition; the statement is vacuous in that case.

The final sentence follows by combining the just-proved status of field j with Theorem 5.1 applied to the other field i . $\square$

Proof of Theorem 5.4. Let

$$F(x, \tau) = \text{BR}(x; \pi, \tau) - x. \quad (144)$$

The strict local sign condition for the two human pre-projection terms means that, in a neighborhood of (x_0, τ_0) , each projection coordinate $[z]_+$ is either the identity map applied to a strictly positive argument or the constant zero map applied to a strictly negative argument. Thus no projection kink is encountered locally, and the projected human best responses are C^1 on the same neighborhood. Since the remaining primitive maps entering BR are C^1 there, F is C^1 there.

At (x_0, τ_0) ,

$$D_x F(x_0, \tau_0) = D_x \text{BR}(x_0; \pi, \tau_0) - I_6 = -(I_6 - D_x \text{BR}(x_0; \pi, \tau_0)). \quad (145)$$

The assumed invertibility of $I_6 - D_x \text{BR}(x_0; \pi, \tau_0)$ is therefore equivalent to invertibility of $D_x F(x_0, \tau_0)$. The implicit function theorem gives an open interval I around τ_0 and a unique C^1 function $x^* : I \rightarrow \mathbb{R}^6$ such that

$$F(x^*(\tau), \tau) = 0, \quad x^*(\tau_0) = x_0. \quad (146)$$

The identity $F(x^*(\tau), \tau) = 0$ is the fixed-point identity $x^*(\tau) = \text{BR}(x^*(\tau); \pi, \tau)$.

Differentiate this identity:

$$D_x \text{BR}(x^*(\tau); \pi, \tau) \frac{dx^*}{d\tau} + \partial_\tau \text{BR}(x^*(\tau); \pi, \tau) - \frac{dx^*}{d\tau} = 0. \quad (147)$$

Rearranging gives

$$(I_6 - D_x \text{BR}(x^*(\tau); \pi, \tau)) \frac{dx^*}{d\tau} = \partial_\tau \text{BR}(x^*(\tau); \pi, \tau), \quad (148)$$

and inversion yields (46). Finally, the reduced welfare is the composition

$$\mathcal{W}(\tau) = W(\pi; x^*(\tau), \tau), \quad (149)$$

so (47) is the ordinary chain rule. $\square$

Proof of Proposition 5.5. Holding the state $x = (v, h, \theta)$ fixed, the common target ω enters welfare only through

$$m_\ell(r_\ell) = \exp\{-\kappa_\ell(r_\ell - \omega)^2\}. \quad (150)$$

Since r_ℓ is held fixed under the partial derivative,

$$\partial_\omega m_\ell(r_\ell) = 2\kappa_\ell(r_\ell - \omega)m_\ell(r_\ell). \quad (151)$$

Therefore

$$\begin{aligned} \partial_\omega W(\pi; x, \omega) &= \sum_{k=1}^2 \sum_{\ell=1}^2 v_k \pi_{k\ell} \partial_\omega m_\ell(r_\ell) Q_{k\ell}(h, \theta) \\ &= \sum_{\ell=1}^2 2\kappa_\ell(r_\ell - \omega) m_\ell(r_\ell) \sum_{k=1}^2 v_k \pi_{k\ell} Q_{k\ell}(h, \theta) \\ &= \sum_{\ell=1}^2 2\kappa_\ell(r_\ell - \omega) m_\ell(r_\ell) B_\ell(x). \end{aligned} \quad (152)$$

Applying the welfare sensitivity formula (47) from Theorem 5.4 gives

$$\frac{dW}{d\omega} = \partial_\omega W(\pi; x^*(\omega), \omega) + D_x W(\pi; x^*(\omega), \omega) \frac{dx^*}{d\omega}. \quad (153)$$

Substituting (46) with $\tau = \omega$ yields the displayed formula (49).

If $r_1^* = r_2^* = \omega$, then $\partial_\omega m_\ell(r_\ell^*) = 0$ for both fields. Hence $\partial_\omega s = 0$. The human and AI best responses do not depend directly on ω ; the viewer best response depends on ω only through s . Thus $\partial_\omega \text{BR} = 0$. Equation (46) gives $dx^*/d\omega = 0$, and the direct derivative $\partial_\omega W$ also vanishes. Therefore $dW/d\omega = 0$ at that equilibrium. $\square$

Proof of Proposition 5.6. For $\eta \in (0, 1]$,

$$g_\eta(x; \pi) = x \iff (1 - \eta)x + \eta \text{BR}(x; \pi) = x \iff \eta\{\text{BR}(x; \pi) - x\} = 0. \quad (154)$$

Since $\eta > 0$, the last equation is equivalent to $\text{BR}(x; \pi) = x$. Hence changing η cannot change the fixed-point set. $\square$

Proof of Theorem 5.7. By Proposition 5.6, all common- η dynamics have the same fixed-point set as the best-response equation $\text{BR}(x; \pi) = x$. It remains to prove that the contraction condition holds for each η and to compute its comparison spectral radius.

The common- η comparison matrix is

$$M_\eta = (1 - \eta)I_3 + \eta C. \quad (155)$$

Because I_3 commutes with C , every eigenvalue of M_η has the form

$$\mu = 1 - \eta + \eta\nu \quad (156)$$

for some eigenvalue ν of C . The matrix C is nonnegative. Perron-Frobenius theory gives a real nonnegative eigenvalue $\rho(C)$, and all other eigenvalues ν satisfy $|\nu| \leq \rho(C)$. Since $\rho(C) < 1$ and $\eta \in (0, 1]$,

$$|1 - \eta + \eta\nu| \leq 1 - \eta + \eta|\nu| \leq 1 - \eta + \eta\rho(C) < 1. \quad (157)$$

For the Perron eigenvalue $\nu = \rho(C)$, the corresponding eigenvalue of M_η is the positive real number

$$1 - \eta + \eta\rho(C) = 1 - \eta\{1 - \rho(C)\}. \quad (158)$$

Therefore

$$\rho(M_\eta) = 1 - \eta\{1 - \rho(C)\} < 1. \quad (159)$$

Theorem 4.6 then gives a unique steady-state equilibrium and global geometric convergence for every $\eta \in (0, 1]$. The displayed expression is strictly decreasing in η because $\rho(C) < 1$. $\square$

Proof of Theorem 5.8. The relaxed map is

$$g_\eta(x; \pi) = (1 - \eta)x + \eta \text{BR}(x; \pi). \quad (160)$$

Differentiating at x^* gives

$$D_x g_\eta(x^*; \pi) = (1 - \eta)I + \eta D_x \text{BR}(x^*; \pi) = (1 - \eta)I + \eta J. \quad (161)$$

If $Jq = \nu q$ for a nonzero vector q , then

$$D_x g_\eta(x^*; \pi)q = (1 - \eta + \eta\nu)q. \quad (162)$$

Thus the linearized multipliers are $1 - \eta + \eta\nu$ as ν ranges over the eigenvalues of J .

A finite-dimensional discrete linear system is Schur stable if and only if every eigenvalue of its transition matrix has modulus strictly smaller than one. This gives the stated exact criterion for the linearized dynamic. When the criterion holds, choose any α strictly between the spectral radius of $D_x g_\eta(x^*; \pi)$ and one. There exists an equivalent vector norm for which the induced operator norm of $D_x g_\eta(x^*; \pi)$ is less than α . Since g_η is C^1 , continuity of the derivative implies that, in a sufficiently small neighborhood of x^* , the nonlinear map is Lipschitz with constant less than one in that norm. Hence x^* is locally asymptotically stable. If a multiplier has modulus larger than one, the linearized system expands along the corresponding generalized eigenspace, so the fixed point is linearly unstable.

For a real eigenvalue $\nu < -1$, the corresponding multiplier is

$$\mu_\eta = 1 - \eta + \eta\nu = 1 - \eta(1 - \nu). \quad (163)$$

Solving $\mu_\eta = -1$ gives

$$\eta_{\text{flip}} = \frac{2}{1 - \nu}. \quad (164)$$

Because $\nu < -1$, we have $1 - \nu > 2$, so $\eta_{\text{flip}} \in (0, 1)$. For $\eta < \eta_{\text{flip}}$, this real multiplier lies above -1 ; for $\eta > \eta_{\text{flip}}$, it lies below -1 and has modulus larger than one. This proves the flip-threshold statement. $\square$

A.4 Proofs for the one-field benchmark

Proof of Proposition 6.1. If $\varrho = 1$, then $e^A = (1 - \varrho)v = 0$ for every state. At any equilibrium,

$$\theta^* = \bar{\theta}(e^{A,*}) = \bar{\theta}(0) = 0, \quad (165)$$

by the assumption $\bar{\theta}(0) = 0$.

If $\varrho = 0$, then $e^H = \varrho v = 0$ for every state. At any equilibrium, the human pre-projection term is

$$\frac{\bar{h}(0)}{1 + \xi\theta^*} - c. \quad (166)$$

Since $1 + \xi\theta^* \geq 1$ and $\bar{h}(0) \leq c$,

$$\frac{\bar{h}(0)}{1 + \xi\theta^*} \leq \bar{h}(0) \leq c. \quad (167)$$

Thus the pre-projection term is nonpositive, and the dropout condition gives $h^* = 0$. $\square$

Proof of Theorem 6.2. Let (v^*, h^*, θ^*) be any one-field equilibrium. The satisfaction bound gives $s^* \in [-Q_{\max}, Q_{\max}]$, so

$$\underline{v} \leq v^* = \bar{\lambda}(s^*) \leq \bar{v}. \quad (168)$$

For the mixed-regime certificate, suppose

$$\bar{\theta}((1 - \varrho)\underline{v}) > 0, \quad c < \frac{\bar{h}(\varrho\underline{v})}{1 + \xi A_{\max}}. \quad (169)$$

Since $\varrho \in (0, 1)$ and $\bar{\theta}$ is nondecreasing,

$$\theta^* = \bar{\theta}((1 - \varrho)v^*) \geq \bar{\theta}((1 - \varrho)\underline{v}) > 0. \quad (170)$$

Also $\theta^* \leq A_{\max}$ and $\bar{h}$ is nondecreasing, so

$$\frac{\bar{h}(\varrho v^*)}{1 + \xi \theta^*} \geq \frac{\bar{h}(\varrho \underline{v})}{1 + \xi A_{\max}} > c. \quad (171)$$

Therefore the human pre-projection term in (62) is strictly positive, and $h^* > 0$. Thus every equilibrium is mixed.

For the extinction certificate, suppose

$$c \geq \sup_{z \in [0, \bar{v}]} \bar{h}(\varrho z). \quad (172)$$

Since $v^* \in [0, \bar{v}]$,

$$\bar{h}(\varrho v^*) \leq c. \quad (173)$$

Because $1 + \xi \theta^* \geq 1$,

$$\bar{h}(\varrho v^*) \leq c \leq c(1 + \xi \theta^*). \quad (174)$$

By Proposition 4.3, $h^* = 0$. $\square$

Proof of Corollary 6.3. Because $0 < \underline{v} \leq \bar{v}$, all finite displayed ratios are well defined; when an inverse value is $+\infty$, the corresponding mixed-region inequality cannot hold and the extinction inequality is interpreted in the usual extended-real sense. We use two elementary properties of the generalized inverse of a continuous strictly increasing function. If $\bar{h}^{\leftarrow}(y) < \infty$ and $e > \bar{h}^{\leftarrow}(y)$, then $\bar{h}(e) > y$. If $y \geq \bar{h}(0)$ and $e \leq \bar{h}^{\leftarrow}(y)$, then $\bar{h}(e) \leq y$; when $\bar{h}^{\leftarrow}(y) = +\infty$, the same conclusion holds because y is at least the supremum of $\bar{h}$.

For the mixed-region certificate, assume

$$\varrho > \frac{\bar{h}^{\leftarrow}(c(1 + \xi A_{\max}))}{\underline{v}}. \quad (175)$$

Then

$$\varrho \underline{v} > \bar{h}^{\leftarrow}(c(1 + \xi A_{\max})), \quad (176)$$

so

$$\bar{h}(\varrho \underline{v}) > c(1 + \xi A_{\max}), \quad (177)$$

which is equivalent to (65). The second displayed condition in the corollary gives

$$(1 - \varrho)\underline{v} > e_A^0. \quad (178)$$

Since $\bar{\theta}(e) > 0$ for $e > e_A^0$, this implies

$$\bar{\theta}((1 - \varrho)\underline{v}) > 0. \quad (179)$$

Theorem 6.2 then proves that every equilibrium is mixed.

For the human-extinction certificate, assume $c \geq \bar{h}(0)$ and

$$\varrho \leq \min \left\{ 1, \frac{\bar{h}^{\leftarrow}(c)}{\bar{v}} \right\}. \quad (180)$$

For every $z \in [0, \bar{v}]$,

$$\varrho z \leq \varrho \bar{v} \leq \bar{h}^{\leftarrow}(c). \quad (181)$$

By the inverse property stated above, $\bar{h}(\varrho z) \leq c$. Hence

$$\sup_{z \in [0, \bar{v}]} \bar{h}(\varrho z) \leq c, \quad (182)$$

and Theorem 6.2 gives $h^* = 0$ at every equilibrium. $\square$

Proof of Proposition 6.4. Fix $v > 0$ and $h > 0$. Let

$$D(\theta) = h + \theta + 2\epsilon_r. \quad (183)$$

Then

$$r(h, \theta) = \frac{h + \epsilon_r}{D(\theta)}, \quad \frac{\partial r}{\partial \theta}(h, \theta) = -\frac{h + \epsilon_r}{D(\theta)^2}. \quad (184)$$

Also

$$m'(r) = -2\kappa(r - \omega)m(r), \quad \frac{\partial q}{\partial \theta}(h, \theta) = (1 - \varrho)(\tilde{f}^A)'(\theta). \quad (185)$$

Since $\Theta_h(\theta) = vm(r(h, \theta))q(h, \theta)$, the chain rule gives

$$\begin{aligned} \Theta'_h(\theta) &= v \left\{ m'(r) \frac{\partial r}{\partial \theta} q + m(r) \frac{\partial q}{\partial \theta} \right\} \\ &= vm(r(h, \theta)) \left[(1 - \varrho)(\tilde{f}^A)'(\theta) + 2\kappa(r(h, \theta) - \omega) \frac{h + \epsilon_r}{(h + \theta + 2\epsilon_r)^2} q(h, \theta) \right], \end{aligned} \quad (186)$$

which is (69).

The equation $r(h, \theta) = \omega$ is equivalent to

$$\frac{h + \epsilon_r}{h + \theta + 2\epsilon_r} = \omega \iff \theta = \frac{h + \epsilon_r}{\omega} - h - 2\epsilon_r = \theta_\omega. \quad (187)$$

Because $r(h, \theta)$ is strictly decreasing in θ , the condition $\theta \leq \theta_\omega$ is equivalent to $r(h, \theta) \geq \omega$. On $[0, \theta_\omega]$, the intrinsic derivative term $(1 - \varrho)(\tilde{f}^A)'(\theta)$ is nonnegative by assumption, and the mixture term is nonnegative because $r(h, \theta) - \omega \geq 0$, $q(h, \theta) > 0$, and all remaining factors are positive. Thus $\Theta'_h(\theta) \geq 0$ on $[0, \theta_\omega]$, so Θ_h is increasing on that interval.

Now assume there are $\theta_+ \in [\theta_\omega, \bar{\theta})$ and $\bar{\theta} > \theta_\omega$ such that $\Theta'_h(\theta_+) > 0$ and the negative mixture term dominates the intrinsic derivative at $\bar{\theta}$. The domination condition is equivalent to $\Theta'_h(\bar{\theta}) < 0$ because $vm(r) > 0$. The function Θ_h is continuous on $[\theta_+, \bar{\theta}]$ and differentiable on its interior. Since the derivative is positive at θ_+ and negative at $\bar{\theta}$, the maximum of Θ_h on a sufficiently small closed interval containing the positive-slope point and ending at $\bar{\theta}$ cannot occur only at the right endpoint, and it cannot occur only at the left endpoint after moving slightly to the right from a point with positive derivative. Hence a maximizer occurs at some interior point $\theta^* \in (\theta_+, \bar{\theta})$. By Fermat's rule, $\Theta'_h(\theta^*) = 0$, and this point is a local maximizer. $\square$

B Additional experimental figures

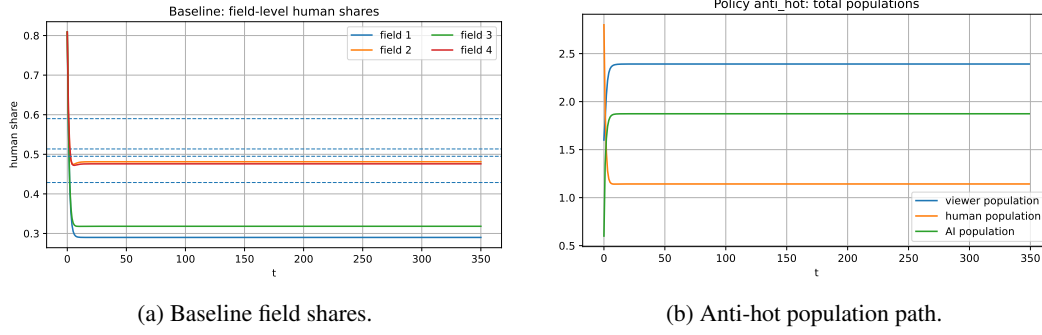


Figure 6: Additional baseline and policy diagnostics.

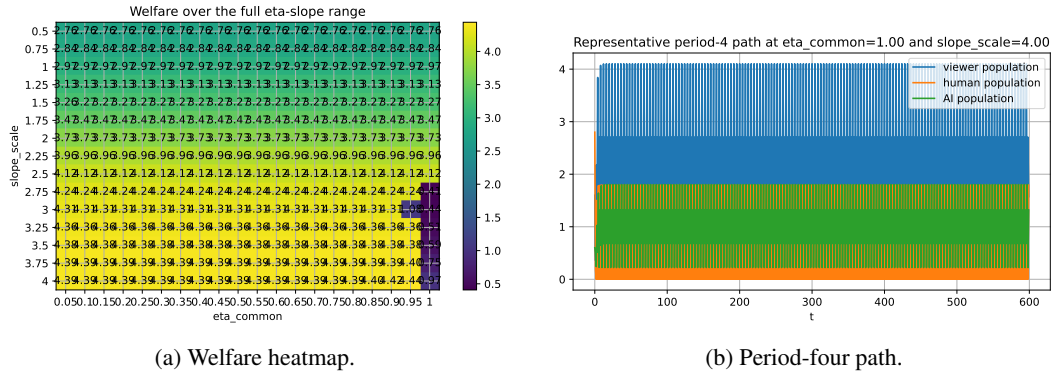


Figure 7: High relaxation and high slope produce persistent nonconvergence.

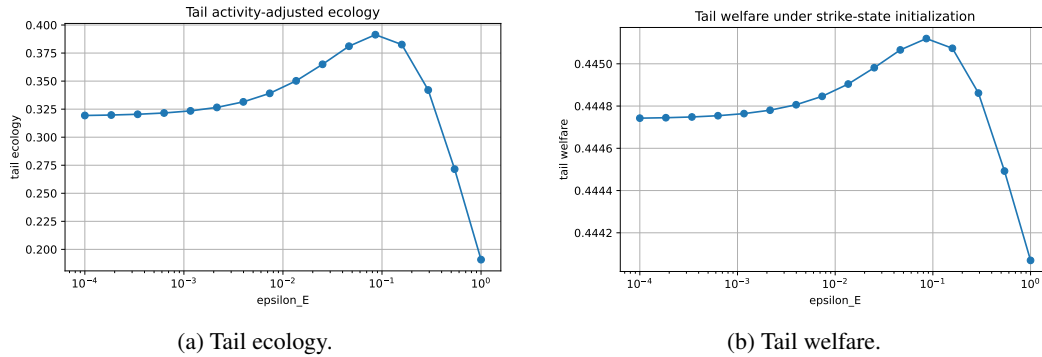
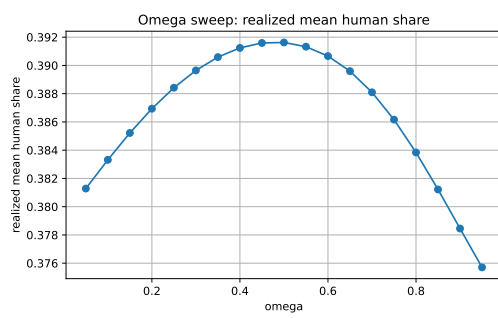
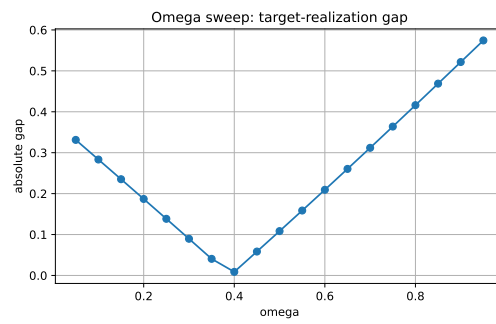


Figure 8: Diagnostics for the ecology smoothing parameter ϵ_E .



(a) Realized human share.



(b) Target-realization gap.

Figure 9: Target-share sweep diagnostics.