Emergent Spatial Traps in the Coevolutionary Commons: Human-AI Cooperation under Environmental Feedback

Ivana Malčić, Debraj Roy, and Luka Waronig

Computational Science, University of Amsterdam,
Science Park 904, 1098 XH Amsterdam, The Netherlands.
ivana.malcic@student.uva.nl, {l.waronig, d.roy}@uva.nl

Abstract. The sustainability of shared resources is increasingly challenged by the rapid expansion of AI infrastructure, whose electricity and water demands intensify pressure on already stressed regions. While coevolutionary game theory shows how environmental feedback shapes cooperation, little is known about human–AI alignment in shared resource systems. We develop a spatial agent-based model in which humans and AI compete over a water commons with environment-dependent payoffs linking local consumption to ecological regeneration. Our findings demonstrate that sustainability in spatial human–AI commons is an emergent property of feedback-mediated incentive alignment rather than resource abundance. Increasing capacity alone can weaken cooperation by allowing extraction pressure to diffuse across the grid, generating spatial traps in which cooperative clusters sustain a stressed environment. Only when environmental responsiveness tightly couples extraction to local consequence can stable interior equilibria or damped oscillations arise. These findings recast alignment in socio-technical systems as a problem of spatially structured feedback design. Alignment is then less a question of intent than of how tightly systems can bind action to consequence across the decision-making space.

Keywords: systems, AI, alignment, cooperation, sustainability, environmental feedback, nonlinear dynamics

1 Introduction

Long-term sustainability with respect to technology depends on how shared natural resources are managed. When individual incentives favor overuse, collective systems can collapse into what is traditionally described as a tragedy of the commons [6], where short-term gains outweigh long-term preservation incentives [13]. The challenge of fostering cooperation in such systems has long been examined, most notably in Axelrod's [14] demonstration that cooperation can emerge without central intervention under repeated interactions. However, this tradition largely treats the environment as static, ignoring environmental feedback. Weitz et al. [19] showed that commons dilemmas cannot be understood independently of the environment they act upon. In particular, they showed that when

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

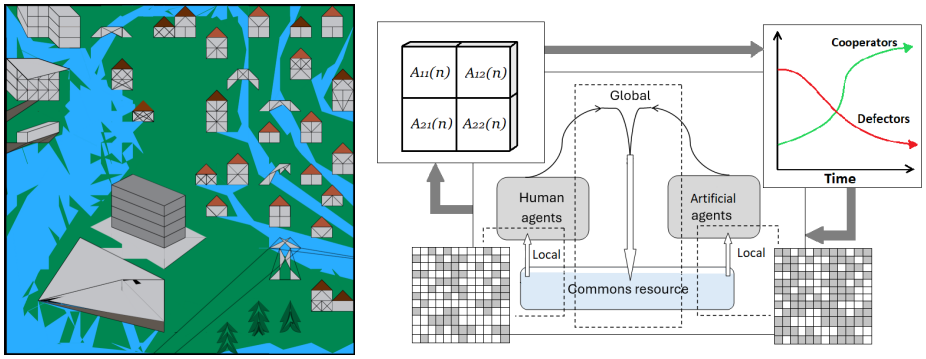
payoff structures depend explicitly on environmental state, strategic incentives and the environment's resource become tightly coupled, producing nontrivial co-evolutionary outcomes even under fully rational decision-making. Their framework, however, assumes a mean-field, well-mixed population, omitting spatial structure and agent heterogeneity.

This paper presents a spatial agent-based extension of the environment-dependent framework of Weitz et al. [19]. By coupling resource-dependent payoff matrices directly to agent decision-making, the model reveals how rational strategies co-evolve with the environmental state, producing stable and oscillatory commons dynamics across parameter space through strategy-mediating localized feedback. To demonstrate this framework in a contemporary and policy-relevant setting, we base the model on a common-pool resource system where human settlements and AI infrastructure share freshwater reserves. The two agent types seek to sustain the shared system and mitigate collapses by switching strategies. Recent work has documented the rapidly growing water footprint of data centers, which can consume millions of liters per day [11] [20] and are often located in regions already experiencing water stress [16]. Agent-based studies have further suggested that AI-driven industrialization can disrupt household water access and trigger complex social responses [1].

In our agent-based model, AI infrastructure is represented through agents with higher baseline water consumption operating under the same strategic payoff structure as human agents. A defecting agent decides to consume more of the resource for their own gain. This design isolates how environmental feedback reshapes rational incentives, and how asymmetries in resource use shift cooperation thresholds, without introducing additional behavioral assumptions that would obscure the core mechanisms. Individual agents repeatedly choose between cooperation and defection based on payoff matrices that evolve with local environmental conditions, while environmental regeneration and strategic incentives shift endogenously as degradation or recovery alters the payoff landscape. As a result, agents remain locally payoff-responsive at all times, and update strategies according to payoff incentives, subject to stochastic deviation. However they become environmentally aligned only when ecological feedback makes cooperation strategically optimal.

The analysis in the remainder of the paper is guided by three questions relevant to the design and interpretation of socio-environmental systems: (1) How do environmental feedback strength, spatial structure, and asymmetric consumption reshape strategic incentives in common-pool resource dilemmas? (2) Under what conditions do feedback-driven systems converge to a stable cooperative state? (3) How does ecological responsiveness determine when cooperation becomes payoff-maximizing for rational, self-interested agents?

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8



(a) Schematic of the environment area with distributed resources. The illustration depicts a large data center in proximity to a human residential area.

(b) Schematic diagram of agent-based feedback-evolving coevolutionary dynamics; ABM extension of Weitz et al. replicator dynamics with feedback-evolving games, where the frequencies of strategy influences n, which modifies the payoffs, $A(n)$.

Fig. 1: Spatial feedback-evolving structure of the agent-based environment.

2 Model Description

2.1 Overview

Purpose: We develop a spatial agent-based model examining how environmental feedback alters strategic incentives in common-pool resource dilemmas to align with long-term resource sustainability, framing human–AI alignment as an ecological necessity. The model extends Weitz et al.'s coevolutionary game-theoretic framework [19] to incorporate spatial structure, local resource competition, and heterogeneous agents. Human households and AI data center infrastructure compete over a shared water commons, where individual consumption affects local resource replenishment and shapes strategic payoffs.

Entities, state variables, and scales: The model consists of three entity types: human agents representing residential households, AI agents representing abstract units of data-center water demand, and grid cells representing the spatial environment and shared water resource. Space is represented as a two-dimensional lattice, and time is discretized such that one model step corresponds to one day.

Model state is stored at two levels; at the agent level, each agent is characterized by an identifier (`id`), a fixed grid position (`pos`), and a binary strategy variable (`strategy` $\in \{C, D\}$). Additional agent-level variables store intended decisions (`planned_action`), selected local resource targets (`target_pos`), and the local payoff matrix (`game`) defining strategic incentives. At the environmental level, grid cells store the shared resource state, including a normalized water level (`W(x,y)`), a local maximum capacity (`W_max(x,y)`), and a baseline

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

replenishment term (`R_base(x,y)`). Cells also record whether they contain a water resource (`has_water(x,y)`) and the set of agents currently occupying the cell (`occupants(x,y)`). The full spatial configuration is represented by the `grid` object, while temporal progression is managed by the model scheduler (`schedule`). Agent types, spatial positions, and cell capacities are fixed at initialization; strategies and water levels evolve over time. A detailed list of all state variables and their data types is provided in [15] (Table 1 in Supplementary Appendix B).

Process overview and scheduling: Algorithm 1 and Algorithm 2 summarize the full scheduling and update order of the agent-based simulation for agent strategy updates as well as environment´s water updates.

Algorithm 1: WaterToC simulation scheduling

Data: 2D grid, agent set A, parameter set, termination time $T_{\max}$
Result: Time series of cooperation, environment state, and other metrics

$t \leftarrow 0$;
while $t < T_{\max}$ **do**
  foreach *water source* w *on grid* **do**
    compute local cooperation fraction c_w;
    update water level $n_w \leftarrow n_w + f(c_w, \theta)$;
  foreach *agent* $i \in A$ **do**
    select random nearby water source w_i;
    evaluate payoff matrix using n_{w_i};
    choose strategy $s_i \in \{\text{cooperate}, \text{defect}\}$ with deviation rate ϵ;
    store planned action a_i;
  foreach *agent* $i \in A$ **do**
    execute a_i and consume water according to s_i;
  record metrics (cooperation, environment state, ...);
  $t \leftarrow t + 1$;

Algorithm 2: Local water replenishment

foreach *water cell* (x, y) **do**
  $x_{\text{loc}} \leftarrow \frac{\#C}{\max(1, \#\text{agents in Moore}_3)}$;;
  $n \leftarrow W / K_{\max}$;;
  $m \leftarrow \text{clip}(1 + 0.5\, n(1-n)(\theta x_{\text{loc}} - (1 - x_{\text{loc}})), 0.1, 3.0)$;;
  $W' \leftarrow \min(K_{\max},\, W + r_{\text{base}} m)$;

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

2.2 Design concepts

1) Theoretical framework: Game-environment feedback

Formalization of coevolutionary games: Strategy-dependent payoffs are fundamental in game theory, representing incentives that influence outcomes. Evolutionary game theory extends this by modeling changes in strategy frequencies through replicator dynamics, which depend on the population's composition of heterogeneous agents [7,10,14]. Building on this foundation, Weitz et al. [19] introduced evolutionary games with environmental feedback, where replicator dynamics couple with environment-dependent payoffs. Their approach begins with the classical symmetric two-strategy Prisoner's Dilemma game, where agents choose between cooperation (C) and defection (D). A standard payoff matrix for such a game is given by:

$$A = \begin{bmatrix} R & S \\ T & P \end{bmatrix}, \quad \text{e.g.,} \quad \begin{bmatrix} 3 & 0 \\ 5 & 1 \end{bmatrix}, \tag{1}$$

where R is the reward for mutual cooperation, S is the sucker's payoff (when a cooperator meets a defector), T is the temptation to defect (when a defector meets a cooperator), and P is the punishment for mutual defection. Under the standard ordering $T > R > P > S$, mutual defection becomes the equilibrium strategy despite the higher collective benefit of mutual cooperation. This is extended by coupling the payoff matrix itself to an environmental state variable $n \in [0, 1]$, which dynamically evolves based on the behavior of the population [19]. This introduces a feedback loop: agents draw from the environment, which in turn alters the incentives in the payoff matrix. Formally:

$$A(n) = (1 - n) \begin{bmatrix} T & P \\ R & S \end{bmatrix} + n \begin{bmatrix} R & S \\ T & P \end{bmatrix}, \tag{2}$$

where $A(n)$ the effective payoff matrix that the agent experiences. This formulation interpolates between a cooperation-favoring game when the environment is depleted (lower n) and a defection-favoring game when the environment is healthy (higher n). The result is a coevolutionary game dynamic, where both strategic behavior and ecological conditions co-determine the future state of the system.

Therefore, a model with feedback-evolving games with asymmetric payoffs is formalized as:

$$A(n) = (1 - n) \begin{bmatrix} R_0 & S_0 \\ T_0 & P_0 \end{bmatrix} + n \begin{bmatrix} R_1 & S_1 \\ T_1 & P_1 \end{bmatrix}, \quad 0 \leq n \leq 1. \tag{3}$$

If n 0, the first payoff matrix dominates and corresponds to a cooperative regime, i.e. $R_0 > T_0$ and $S_0 > P_0$. This makes mutual cooperation the equilibrium under replete conditions. Conversely, if n 1, the second payoff matrix dominates, leading to defector-dominant regime, with $R_1 < T_1$ and $S_1 < P_1$.

By breaking the symmetry of payoffs in this way, Weitz et al. were able to model more nuanced social dilemmas in which not only the relative ranking

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

of payoffs shifts with the environment, but also their magnitude. This allows for exploration of different dynamical regimes emerging from some other game designs and assumptions. The proposed ABM implements the game under fixed symmetric payoffs as in (2), using values from example (1) because they preserve the relative ordering which is representative of the problem underlying this study.

2) Agent decision-making: In the model, cooperation and defection correspond to qualitatively different resource-use behaviors for both agent types. For human agents, cooperation represents responsible household water use consistent with long-term sustainability, while defection reflects increased extraction in response to perceived competition, intensifying local consumption pressure and reducing availability for others. For AI agents, cooperation corresponds to operating within efficiency and sustainability constraints, whereas defection represents disproportionate resource extraction aimed at maximizing performance without regard for local limits.

3) Environmental dynamics and feedback: Figure 1 illustrates the spatial environment of the model, which consists of a 20×20 grid, with local water resources coupled to agent interactions. Water resources are distributed across the grid according to the water-cell density ρ_w, and each water cell is characterized by a maximum capacity K_{max} and a baseline replenishment rate representing background renewal.

Environmental change is governed by feedback from local agent behavior. Resource regeneration depends on the fraction x of cooperating agents within a Moore neighborhood of radius 3 surrounding each water cell. This feedback captures the idea that cooperative behavior enhances local environmental recovery, while defection increases degradation pressure. In a mean-field approximation, the resulting dynamics can be expressed as a bounded growth process,

$$\frac{d}{dt}n = n(1-n)\left[\theta x - (1-x)\right] \tag{4}$$

where n denotes the normalized local water level, x denotes cooperators, and θ controls the relative strength of cooperative restoration against degradation.

The parameter θ can be described as the ratio between the positive contribution of cooperators and the negative impact of defectors on the environment. For $\theta > 0$, cooperative actions allow the environment to recover efficiently despite ongoing consumption - in this regime, the resource is effectively more renewable, as cooperative behavior can offset or reverse degradation. Conversely, when θ is small cooperative actions contribute weakly to regeneration, defection dominates environmental change, and resource renewal is limited which is why recovery becomes slow or impossible. In its functional form, the equation can be written as $f(x) = \theta x - (1-x)$. This formulation ensures that environmental recovery is strongest at intermediate resource levels and limited near depletion or saturation, consistent with standard ecological regeneration models.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

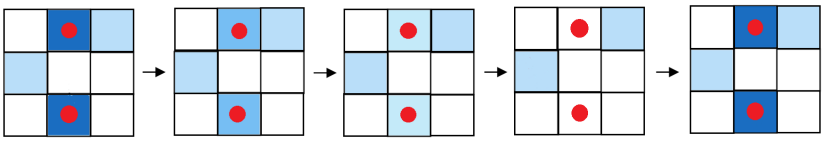


Fig. 2: Individual agent-environment sensing (1. Agents sense local water availability; 2. Agents consume water from cells; 3. Environment resource depletes; 4. Mutual consumption reinforces cooperation; 5. Environment replenishes)

4) Sensing: Agents perceive the resource from their neighboring cells, i.e. 'sense' their environment. Each agent makes decisions based on local information rather than global averages and adapts strategies in response to the immediate water availability in their surroundings. (This process was detailed in Algorithms 1 and 2.) Figure 2 illustrates the local activity at discrete time steps, starting from agents sensing their local environment for available water capacity in a Moore neighborhood with a radius of 1, implemented to preserve locality and impacts felt at the household level.

5) Output metrics: Primary output metrics are: (1) the final cooperation fraction as the proportion of agents (both human and AI) adopting cooperative strategies at the end of each simulation - serving as the primary indicator of system-wide sustainability outcomes and reflects the success of collective action; and (2) cooperation stability, which is measured as the variance in cooperation levels in the final stage of the simulation. Low variance indicates stable regimes, while high variance suggests oscillatory dynamics characteristic of the oscillating tragedy of the commons from literature.

3 Experimental setup

3.1 Model parameterization and initialization

Water cells are placed according to the water-cell density ρ_w, and each cell is characterized by a maximum water capacity K_{max}. The simulation hosts 50 human agents representing household units and 50 AI agents representing water-intensive compute infrastructure.

The water consumption parameters were derived by preserving ratios of real-world statistics of water use. A typical U.S. household's daily water use of approximately 1000 L was taken as the baseline for a cooperating human agent [18], while a defecting strategy was modeled as a 1.5× increase to 1500 L per day. For AI agents, the average daily consumption of a moderate-sized data center (approximately 1,000,000 L) was distributed across 50 AI agents, yielding a cooperative baseline of 20,000 L per agent [8]. Applying the same 1.5× multiplier gives 30,000 L per AI agent under defection. Anchoring the human

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

Table 1: Model parameter descriptives

Parameter	**Symbol**	**Bounds**	**Baseline**
Global parameters			
Grid size	L	$[1, \infty)$	20
Initial human agents	N_h	$[0, \infty)$	50
Initial AI agents	N_{ai}	$[0, \infty)$	50
Max water capacity per cell	K_{max}	$[0, \infty)$	20
Water cell density	ρ_w	$[0, 1]$	0.3
Feedback strength	θ	$(-\infty, \infty)$	3
Strategy deviation rate	ϵ	$[0, 1]$	0.1
Agent-specific parameters			
Human consumption (cooperate)	C_h	$[0, \infty)$	0.1
Human consumption (defect)	D_h	$[0, \infty)$	0.15
AI consumption (cooperate)	C_{ai}	$[0, \infty)$	2.0
AI consumption (defect)	D_{ai}	$[0, \infty)$	3.0
Game payoffs (R, S, T, P)	–	$(-\infty, \infty)$	(3, 0, 5, 1)

cooperative value at 0.1, the remaining parameters were derived proportionally: human defection at 0.15, AI cooperation at 2.0, and AI defection at 3.0. This ratio-based scaling preserves the relative intensity of residential and industrial water use while allowing the model to operate on normalized resource units. Model parameters and their values are summarized in Table 1.

3.2 Experiments

To analyze the model's behavior, a comprehensive parameter sweep was performed. The specific parameter ranges explored for feedback strength (θ), max water capacity (K_{max}), and water cell density (ρ_w) are presented in Table 2. Ultimately 64 parameter combinations were run for 100 iterations with 100 time steps (representing days) per run to ensure robust statistical analysis and address the model's stochasticity (total of 6400 simulations).

Table 2: Parameter values for experimental sweep

Parameter	**Explored parameter space**
Feedback strength (θ)	1.4, 2.0, 5.0, 10.0
Max water capacity (K_{max})	10, 20, 30, 40
Water cell density (ρ_w)	0.2, 0.3, 0.4, 0.5

First, to determine the effects of environmental factors on the final cooperation and environment states, the outcomes of all 100 runs were aggregated for each parameter combination. The mean value of the final cooperation fractions and environment states at the last time step was calculated. These aggregated results were then used to identify trends.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

Second, to identify conditions under which the system stabilizes at an equilibrium, each individual simulation run was post-processed. A run was classified as having reached a stable fixed-point if (i) the computed variance of its cooperation fraction over the final 20% of time steps fell below a set threshold (10^{-3}) and (ii) the time series was confirmed as stationary by an Augmented Dickey-Fuller (ADF) test (p–value < 0.05). Aggregating across stochastic replications allowed us to map parameter regions where more than 50% of runs stabilized around such point. This fixed–point screening is complementary to the derived mean–field model, whose equations predict possible interior equilibria ($x^* > 0, n^* < 1$) and their stability under the assumption of an infinite well–mixed population, while the agent–based analysis checks whether those predicted attractors actually tend to emerge in finite, stochastic, spatially structured simulations. The full derivation of the mean-field system, its linearization and phase-space, are shown in [15] (Supplementary Appendix A, Figs. S1–S2). Moreover, spatial metrics were collected to analyze emergent patterns on the grid. Across runs we compute the number of clusters, overall cooperation fraction, size of the largest cooperative cluster, and the spatial autocorrelation measured by Moran's I test.

Finally, global sensitivity analysis [2] was conducted using two methods: (1) total-order Sobol indices, and (2) the Kolmogorov-Smirnov (KS) test (confirming strong nonlinear interactions guiding the behavior of the model, with the detailed results shown in Figs. S9–S10 of Supplementary Appendix B [15]).

4 Results and Analysis

High cooperation fractions in the population are most favored under conditions of local scarcity and strong environmental feedback. Increasing resource abundance weakens cooperation by delaying the ecological consequences of defection, while sufficiently strong feedback stabilizes behavior through fixed points or oscillatory dynamics. All regimes exhibit strong spatial structure, with cooperation stabilized locally through clustered resource–strategy feedback.

4.1 Evolution of cooperation

After conducting a thorough sweep across water cell density (ρ_w), maximum water capacity ($K_{\max}$), and feedback strength (θ), it was revealed that cooperation levels were highest under conditions of relative scarcity, namely when both ρ_w and $K_{\max}$ were low. In these settings, agents faced strong local competition for limited resources, and defectors were quickly punished as depletion occurred almost immediately. This dynamic strengthened the relative advantage of cooperation, since cooperative behavior more effectively sustained resource availability over repeated interactions.

In contrast, cooperation is lowest when both resource density and capacity are high. Under abundance, defectors can extract resources for longer without immediate depletion, delaying punishment and reducing long-term cooperation. The tragedy of the commons here is stronger: defection pays off for longer before

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

scarcity is felt, lowering the long-term cooperation fraction. Thus, abundance can weaken cooperation rather than reinforce it.

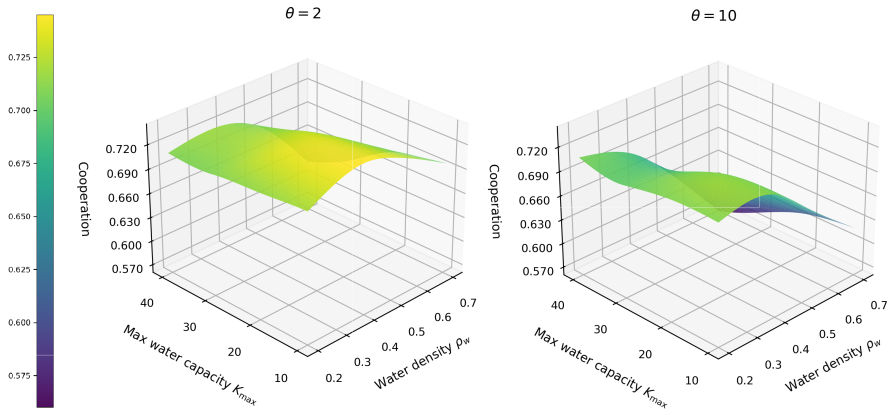


Fig. 3: Comparison of landscape for lower ($\theta = 2$) vs. higher feedback sensitivity ($\theta = 10$). The colour gradient reveals that highest cooperation fractions are reached in scenarios with lower resource capacity and density. Maximizing θ lowers cooperation.

The feedback strength parameter θ played a stabilizing role in this dynamic. As θ increased, the sensitivity of cooperation to variation in $K_{\max}$ diminished, particularly when ρ_w was low. Extended heatmaps and 3-d coupled-system landscape across the full parameter sweep are shown in [15] (Figs. S3–S5 in Supplementary Appendix B). Stronger feedback mechanisms ensured that cooperative actions more directly translated into local replenishment, offsetting the destabilizing effect of larger immediately-available capacities. At $\theta = 10$, cooperation levels remained constant across $K_{\max}$ values when $\rho_w = 0.2$, indicating robust stability under conditions of scarcity. This motivates an analysis of fixed point attractors which nudge the system onto its most stable trajectories across stochastic runs. Together, these results indicate that cooperation is not solely determined by resource abundance or any other single lever, but by the interaction between ecological structure and feedback strength.

4.2 Fixed point attractors

A detailed analysis of the fixed-point attractors, shown in Fig. 4, highlights the characteristic behavior of the system when the variance stabilizes. Stable equilibria occur most frequently under sparse resource conditions ($\rho_w = 0.2$), where the majority of runs converge to interior fixed points with both cooperation and the environment settling to nontrivial values different than the mean field approximation. Across these attractors, the cooperation fraction consistently stabilizes

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

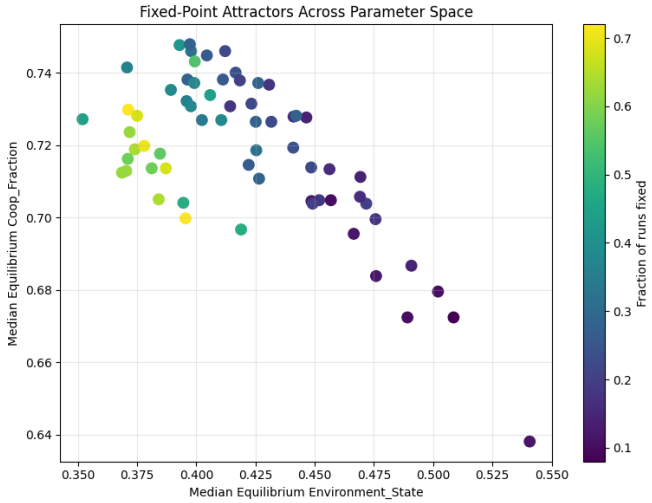


Fig. 4: Scatter plot of fixed-point attractors across the entire parameter space. Each point represents the median equilibrium outcome for a unique parameter combination, plotting the average equilibrium environment state on the x-axis against the equilibrium cooperation fraction on the y-axis.

at relatively high levels ($x^* \approx 0.70$), while the environment remains moderately stressed ($n^* \approx 0.36{-}0.40$). This inverse relationship reveals a core tension of the system: even widespread cooperation cannot restore the commons to abundance, but instead sustains it at an intermediate, pressured state. Cooperation moderates consumption just enough to balance ongoing demand, preventing collapse without enabling full recovery. Detailed statistics for all 64 parameter combinations are reported in [15] (Table 2 in Supplementary Appendix B).

The scatter of points in Fig. 4 demonstrates that the position of this equilibrium is not fixed but shifts with ecological and demographic parameters. In particular, increasing the maximum cell capacity (K_{max}) tends to push the equilibrium towards slightly lower cooperation fractions, despite allowing the environment to stabilize at higher abundance (see Fig. S5 in Supplementary Appendix B [15]). For example, at $\theta = 5.0$, raising K_{max} from 10 to 40 reduces cooperation from 0.720 to 0.706. This suggests that larger potential resource rewards indirectly incentivize more defection, as agents can extract more before depletion becomes a constraint. The interior fixed point corresponds not to homogeneous cooperation, but to a spatially heterogeneous configuration in which cooperative clusters sustain nearby resources while peripheral regions remain chronically depleted.

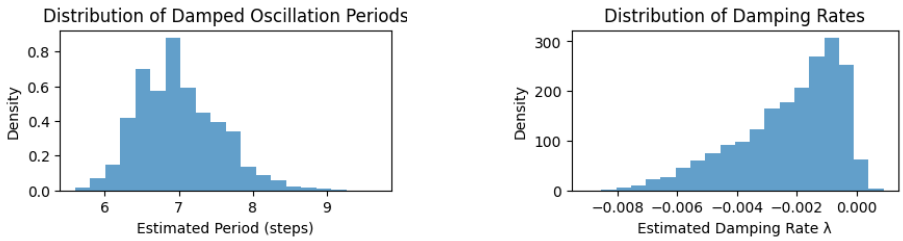
ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

4.3 Oscillatory dynamics

Under certain parameter regimes, the coupled feedback between strategic behavior and environmental state produces oscillatory dynamics rather than convergence to a fixed point. In these regimes, cooperation levels and resource availability evolve periodically without external forcing, forming closed trajectories in phase space characteristic of limit cycles [17]. In the present model, such oscillations arise from delayed feedback between cooperation and environmental regeneration: changes in strategy alter resource levels, but the resulting shift in incentives is not immediately perceived by agents.

When the feedback strength θ is low, improvements in cooperation translate only slowly into environmental recovery. This temporal lag decouples short-term strategic incentives from longer-term ecological benefits, destabilizing fixed-point behavior. As θ increases, feedback becomes sufficiently strong to sustain coordinated oscillations, in which cooperation rises as resources recover and subsequently declines as defection becomes temporarily advantageous. The resulting dynamics reflect an internally generated rhythm driven by the coevolution of strategy and environment, rather than by exogenous perturbations.

To characterize these oscillations quantitatively, two complementary time-series analyses were applied. First, a peak-envelope analysis for water cell densities of 0.2 and 0.4 identified local maxima in the cooperation fraction $C(t)$. Inter-peak intervals yield an average oscillation period of $\overline{T} = 7.0 \pm 0.54$ steps (IQR: 6.6–7.4). Fitting an exponential envelope $A_0 e^{-\lambda t}$ to the peak amplitudes gives a mean damping rate of $\overline{\lambda} = -0.00226 \pm 0.00173$ per step (IQR: -0.00332–-0.00086). The results are summarized in Figures 5a and 5b.



(a) Distribution of damped oscillation periods.

(b) Distribution of damping rates.

Fig. 5: Damped oscillatory behavior in the cooperation–environment dynamics.

Second, a *Morlet-wavelet continuous transform* was applied to six representative parameter combinations ($\theta \in \{2, 5, 10\}$, $K_{\max} \in \{20, 40\}$, and densities $\rho_w = 0.2$ and 0.4). Across all cases, the resulting scalograms (see Fig. S6 in Supplementary Appendix B [15]) exhibit a single, well-defined ridge at frequency $f \approx 0.14$ (1/step), corresponding to a period of approximately seven time steps

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

and confirming the same dominant oscillatory mode. From a spatial perspective, these oscillations correspond to distributed depletion–recovery fronts on the grid, where cooperative and defective regions alternately expand and contract across neighboring locations rather than synchronizing globally. The oscillatory regime therefore reflects spatially localized adjustment processes, reinforcing the role of local ecological feedback in shaping global system behavior.

4.4 Spatially-embedded dynamics

Positive spatial autocorrelation in cooperation (Moran's $I > 0$) indicates that cooperative strategies cluster rather than disperse randomly. These clusters align with regions of higher resource availability, confirming that cooperation is stabilized through local feedback rather than global averaging.

Increasing feedback strength θ increases the size and persistence of the largest cooperative cluster, meaning that stronger feedback enlarges the spatial basin of attraction for cooperation. The complete detailed clustering characteristics that were found are reported in Figs. S7–S8 and Tables S3–S4 in Supplementary Appendix B [15]. Importantly, the emergence of cooperation in this model is inherently spatial. Unlike mean-field formulations, agents here experience environmental feedback at the same spatial scale at which extraction occurs. Cooperation therefore stabilizes first in localized regions where collective restraint preserves nearby resources, while defection is selectively punished through rapid local depletion. The resulting dynamics do not converge to a homogeneous cooperative state, but to a heterogeneous spatial mosaic composed of strategic clusters embedded within stressed environments.

The oscillatory regimes are spatially distributed depletion–recovery cycles, where cooperative regions expand and contract over time. This spatial coupling explains why increased resource abundance destabilizes cooperation: higher density and capacity weaken the link between local action and local consequence, allowing defectors to evade depletion by shifting pressure across the grid. Cooperation in this configuration thus emerges not from coordination, but from the alignment of strategic incentives with spatially localized ecological feedback.

5 Discussion

This study demonstrated how sustainability in shared-resource systems is shaped less by aggregate abundance and access to capacity, rather more by how local feedback structure creates incentives across the space. Even when global conditions appear favorable, cooperation can be trapped in spatially localized pockets, while surrounding regions experience chronic depletion. These spatial traps arise when agents can shift extraction pressure across the grid faster than local ecological feedback can penalize defection, weakening the link between individual behavior and nearby environmental consequences.

Within cooperative clusters, agents experience a reinforcing loop: cooperative neighborhoods maintain higher local water levels, which in turn keep cooperation

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

payoff-maximizing. Yet this advantage does not translate into system-wide sustainability. Defectors continue to exploit peripheral regions, sustaining a configuration in which high cooperation coexists with persistent environmental stress. This 'managed scarcity' equilibrium is kept just above collapse by cooperative pockets, never transitioning to a high-cooperation, high-abundance regime. Oscillatory dynamics follow the same spatial logic: depletion-recovery cycles propagate as moving fronts where cooperative clusters expand, exhaust their local buffer, and retreat as defection temporarily becomes advantageous. Increasing capacity or density does not simply lift the resource baseline; it redistributes extraction pressure across the grid, deepening traps by allowing defectors to avoid localized consequences.

Focusing on such spatial traps allows us to reframe governance implications. Interventions which only increase aggregate capacity or set uniform rules, risk entrenching patterns where some segments act as buffers that absorb stress for others. In that sense, heterogeneity can act as a stabilizing endogenous force. More promising would be mechanisms that tighten the coupling between local use and local consequences; such as spatially differentiated pricing, locally adaptive caps, or infrastructure siting rules that limit expansion into residential areas. In modeling terms, this to strengthening or rescaling environmental feedback so that cooperative behavior is rewarded, while defection penalized, at the same spatial scale where decisions are made.

The findings are also consistent with broader studies that emphasize the importance of a system's institutional fit between resource characteristics and governance mechanisms [5,12]. Rather than targeting single levers such as capacity expansion, effective governance must account for the coupled evolution of behavior and environment. Related studies similarly argue that commons dilemmas are most effectively addressed when institutions directly link individual actions to ecological outcomes [4,3], a principle that emerges endogenously in our model through environment-dependent payoffs. Future research can extend this framework by studying other asymmetric agent configurations and even model explicitly institutional heterogeneity across the space (e.g. centralized vs decentralized management) to study conditions for alleviation or reinforcement of the spatial traps. Alignment, therefore, is less a question of intent than of how tightly systems bind action to outcome. In a spatial resource commons, long-term sustainability can be achieved by neither prioritizing nor eliminating self-interest, rather by ensuring that no action escapes its own ecological footprint.

Supplementary information and code availability: Supplementary Information can be found at [15]. All code is public [9] and can be used to reproduce and verify the results of this study.

References

1. Artificial intelligence, human agency and social decision-making in water management systems. TISSS Lab, Johannes Gutenberg University (2024)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8

2. Bazyleva, V., Garibay, V.M., Roy, D.: Trajectory-based global sensitivity analysis in multiscale models. Scientific Reports **14**(1), 13902 (2024)
3. Borowski-Maaser, I., G.M.F.N.P.M.R.A.H., Boogaard, F.: Watercog: Evidence on how the use of tools, knowledge, and process design can improve water co-governance. Water **13**(9), 1206 (2021). https://doi.org/10.3390/w13091206, https://doi.org/10.3390/w13091206
4. Druzin, Bryan, H.: The parched earth of cooperation: How to solve the tragedy of the commons in international environmental governance. Duke Journal of Comparative & International Law **27**(1), 73–105 (2016). https://doi.org/10.2139/ssrn.2829141, https://scholarship.law.duke.edu/djcil/vol27/iss1/3
5. Dupont, C., Roy, D.: Emergent poverty traps at multiple levels impede social mobility. Humanities and Social Sciences Communications **12**(1), 1777 (2025)
6. Hardin, G.: The tragedy of the commons. Science **162**(3859), 1243–1248 (1968). https://doi.org/10.1126/science.162.3859.1243, http://www.jstor.org/stable/1724745
7. J., M.S.: Evolution and the theory of games. Cambridge University Press, Cambridge UK (1982)
8. Kreutzer, M.: The hidden costs of the digital age: A case study on server impacts on the environment. Medium (2025)
9. Malčić, I., Waronig, L.: Water-toc agent-based model (2025), https://github.com/luka-waronig/WaterToC
10. Mengesha, I., Roy, D.: Evolutionary game selection leads to emergent inequality. In: International Conference on Computational Science. pp. 284–297. Springer (2025)
11. Mytton, D.: Data centre water consumption. npj Clean Water **4**(1), 11 (2021)
12. Olivier, T., Vallury, S.: Institutional fit and policy design in water governance: Nebraska's natural resources districts. Policy Studies Journal **52**(4), 809–832 (2024). https://doi.org/10.1111/psj.12550, https://doi.org/10.1111/psj.12550
13. Ostrom, E.: Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press, Cambridge, UK (1990)
14. R., A., W.D., H.: The evolution of cooperation. Science pp. 1390–1396 (1981). https://doi.org/10.1126/science.7466396
15. Roy, D., Malcic, I., Waronig, L.: Supplementary information: Emergent spatial traps in the coevolutionary commons: Human-ai cooperation under environmental feedback (2026). https://doi.org/10.5281/zenodo.18713008
16. Skidmore, Z.: Ai data center growth deepens water security concerns in high-stress states – report. Data Center Dynamics (May 2025), https://www.datacenterdynamics.com/en/news/ai-data-center-growth-deepens-water-security-concerns-in-high-stress-states-report/
17. Strogatz, S.H.: Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. Westview Press, 2nd edn. (2015)
18. U.S. Environmental Protection Agency: How we use water (2024), https://www.epa.gov/watersense/how-we-use-water, "The average American family uses more than 300 gallons of water per day at home."
19. Weitz, J. S., E.C.P.K.B.S.P., Ratcliff, W.C.: An oscillating tragedy of the commons in replicator dynamics with game-environment feedback. Proceedings of the National Academy of Sciences **113**(47), 7518–7525 (2016)
20. Yañez-Barnuevo, M.: Data centers and water consumption (June 2025), https://www.eesi.org/articles/view/data-centers-and-water-consumption

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_8