Nonlinear Science 9 (2026) 100224

Contents lists available at ScienceDirect

Nonlinear Science

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

Research paper

From interaction density to evolutionary time: The mathematical dynamics behind cooperation in large evolutionary populations

Óscar De los Reyes-Marín [a,*], Iria Paz Gil [b], Jose Torres-Pruñonosa [c], Raúl Gómez-Martínez [b]

[a] Doctoral Researcher in Behavioral Finance Rey Juan Carlos University (URJC), Calle Tulipán, s/n, 28933, Móstoles, Madrid, Spain
[b] Rey Juan Carlos University (URJC), Calle Tulipán, s/n, 28933, Móstoles, Madrid, Spain
[c] International University of La Rioja (UNIR), Av. de la Paz, 137, 26006, Logroño, La Rioja, Spain

ARTICLE INFO

JEL classification:
C73
C63
C62
C61

Keywords:
Interaction density
Evolutionary time
Evolutionary game theory
Fermi imitation
Finite-size effects
Mean-field dynamics

ABSTRACT

Interaction density was recently identified as the dominant predictor of cooperation in large evolutionary populations governed by continuous cooperation propensities, random encounters, Fermi imitation, and mutation. Here, we derive the mathematical dynamics underlying that computational phenomenon. We show first that, in the audited sequential Fermi process, interaction density does not modify the microscopic transition kernel. Instead, it exactly rescales evolutionary exposure through the natural interaction clock $\tau = 2gI/N$, where g is the number of generations, I the number of encounters per generation, and N the population size. Consequently, for positive mutation, the stationary finite-population distribution is independent of I, and the previously observed interaction-density landscape is fundamentally a finite-horizon relaxation phenomenon rather than an interaction-induced stationary bifurcation.

The Fermi kernel admits an exact reduction in which selection depends on payoffs only through $\beta(T-S)$. We derive the exact finite-population conditional drift of mean cooperation and show that the mean alone does not close the dynamics. In the mutation-free mean-field limit with uniform initialization, the distribution equation admits the explicit solution $C(\tau) = 1/(a\tau) - 1/(e^{a\tau}-1)$, with $a = \tanh\left[\beta(T-S)/2\right]$, providing a closed-form explanation for the observed cooperation plateau and subsequent decline.

For finite populations, simulations and exact drift identities reveal a transient slowdown relative to mean-field evolution. This slowdown is not primarily caused by direct $1/N$ sampling corrections, but by finite-size distortions of the strategy distribution. A moment hierarchy links variance production to skewness and skewness dynamics to the fourth cumulant. Because low-order moment closures fail dynamically, we construct a reduced distributional representation. The mean-field trajectory is highly compressible, and a low-dimensional latent flow predicts later dynamics outside its fitting window.

Finally, prospectively generated simulations with frozen models and prespecified criteria show that finite populations move in nearly the same latent direction as the mean-field flow and that independently estimated effective evolutionary speeds agree within the prespecified tolerance. Strict monotonic decay of the transverse finite-size component with population size remains unresolved. Together, these results provide a mathematical account of the interaction-density dependence previously observed computationally and connect microscopic imitation,

* Corresponding author.
E-mail addresses: o.delosreyes.2016@alumnos.urjc.es (Ó.D.l. Reyes-Marín), iria.paz@urjc.es (I.P. Gil), jose.torresprunonosa@unir.net (J. Torres-Pruñonosa), raul.gomez.martinez@urjc.es (R. Gómez-Martínez).

https://doi.org/10.1016/j.nls.2026.100224
Received 16 September 2026; Received in revised form 30 September 2026; Accepted 1 October 2026
Available online 8 October 2026
3050-5178/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

evolutionary time, finite-size distributional effects, and macroscopic cooperation within a unified dynamical framework.

1. Introduction

Cooperation among self-interested agents remains one of the central problems of evolutionary game theory, behavioral economics, statistical physics, and complex-systems science. Classical explanations typically rely on mechanisms such as repeated interaction, reciprocity, reputation, punishment, assortment, network structure, or group selection. Yet cooperation can also display non-trivial collective patterns in populations where agents possess none of these stabilizing mechanisms and interact only through local stochastic encounters. Understanding which microscopic features are sufficient to generate such macroscopic regularities is therefore essential not only for explaining cooperation itself, but also for determining when an apparent structural effect is instead the transient manifestation of an underlying dynamical process.

In our previous study, *Interaction Density Drives Cooperation in Large Evolutionary Populations: Evidence from Large-Scale Agent-Based Simulations* [1], we investigated this problem using large populations of agents playing a Prisoner's Dilemma under deliberately minimal behavioral assumptions. Agents were characterized by continuous cooperation propensities, interacted through random pairwise encounters, and adapted through Fermi imitation with mutation. They possessed no memory of previous encounters, no reputation, no persistent partners, and no stable reciprocity mechanism. Large-scale computational exploration revealed a highly structured cooperation landscape in which interaction density emerged as the dominant predictor of the observed macroscopic cooperation level. In particular, cooperation remained comparatively high at low and intermediate interaction densities and declined progressively when interaction exposure became sufficiently large. This behavior was summarized empirically by a relationship of the form

$$C = F(I, \mu),$$

where C denotes observed cooperation, I interaction density, and μ mutation intensity.

That result established a robust computational phenomenon, but it left its governing dynamics unresolved. In particular, it remained unclear whether interaction density altered the stationary evolutionary state of the system, whether it generated a genuine structural transition, or whether it instead controlled the rate at which populations moved through a common evolutionary process. These interpretations are fundamentally different. A true interaction-induced transition would imply that changing I modifies the effective evolutionary regime itself. By contrast, a temporal interpretation would imply that populations observed after the same number of nominal generations may actually have experienced very different amounts of evolutionary exposure. Distinguishing these possibilities requires moving beyond equilibrium summaries and machine-learning importance measures toward a mathematical derivation from the microscopic transition rules.

The present study addresses that problem directly. We show that, for the audited sequential Fermi process underlying the revised simulations, interaction density has a simple but consequential mathematical role: it determines the number of microscopic encounters executed per generation but does not enter the one-encounter transition rule itself. If P denotes the Markov transition operator associated with one encounter, then one generation containing I encounters is governed by

$$P_I = P^I.$$

Consequently, after g generations,

$$\mathcal{L}(X_g) = \mathcal{L}(X_0)\, P^{gI},$$

so that the natural evolutionary clock is not the generation index g, but rather the accumulated exposure

$$\tau = \frac{2gI}{N},$$

where N is population size and the factor $2/N$ converts encounters into expected participations per agent. This identity implies that, at equal accumulated numbers of encounters, different choices of I generate the same stochastic process up to a regrouping of encounters into generations. Thus, within this model, interaction density is fundamentally a clock parameter. This clock equivalence concerns sequential encounter-level updating; it does not apply automatically to protocols that accumulate payoffs before a selection event.

This reinterpretation has immediate consequences for the interaction-density landscape reported previously. When simulations are compared after the same number of generations, larger values of I correspond to substantially larger values of τ. The system is therefore being observed at progressively later points along its evolutionary relaxation. The decline in cooperation associated with increasing I can consequently arise without any interaction-induced stationary bifurcation. For positive mutation and finite selection intensity, the stationary finite-population process is common to all $I > 0$; interaction density determines how quickly that regime is approached in generational time, not which stationary regime exists. The previously observed dependence on I is therefore a genuine finite-horizon property of the model, but its mathematical origin is temporal rather than equilibrium-selecting.

The microscopic Fermi rule also admits an exact reduction. Let x and y denote the cooperation propensities of two interacting agents and let β denote selection intensity. Averaging over the four possible Prisoner's Dilemma action outcomes yields the marginal imitation probability

$$q(x,y) = \frac{1}{2} + \frac{a}{2}(x-y),$$

where

$$a = \tanh\left[\frac{\beta(T-S)}{2}\right].$$

Thus, for the audited update rule, the selective component depends on the payoff matrix only through the composite quantity $\beta(T-S)$; the reward R and punishment P cancel from the pairwise comparison. This reduction substantially simplifies the apparent parameter space and provides the first step toward a tractable macroscopic description.

The dynamics of mean cooperation, however, do not close in the mean alone. For a finite population with cooperation propensities $\boldsymbol{p} = (p_1, \ldots, p_N)$, the exact conditional drift of the population mean $C = N^{-1}\sum_i p_i$ can be written as

$$\frac{\mathbb{E}[\Delta C \mid \boldsymbol{p}]}{2/N} = -ak_N V + \frac{1}{2}B - \frac{a}{2}k_N D,\ k_N = \frac{N}{N-1},$$

where V is the variance of the cooperation propensities and B and D capture mutation and clipping effects. Two populations can therefore have the same C while exhibiting different expected drifts because their internal distributions differ. The empirical relationship $C = F(I,\mu)$ observed at the macroscopic level is consequently not the fundamental dynamical law of the system. The underlying state variable is the distribution of cooperation propensities.

This observation leads naturally to a distributional mean-field formulation. In the large-population limit, the state of the population is represented by a probability distribution $f_\tau(p)$ over continuous cooperation propensities. Its evolution is governed by a nonlinear mutation-selection equation whose first moment is the macroscopic cooperation level

$$C_\tau = \int_0^1 p\, f_\tau(p)\, dp.$$

In the mutation-free case, the mean-field equation can be solved explicitly. For an initially uniform distribution,

$$f_\tau(p) = \frac{f_0(p)\, e^{-a\tau p}}{Z(\tau)},$$

which yields

$$C(\tau) = \frac{1}{a\tau} - \frac{1}{e^{a\tau}-1}.$$

Substituting $\tau = 2gI/N$ provides an explicit analytical relation between cooperation, interaction density, population size, selection intensity, and observation horizon. The expression reproduces the qualitative structure observed computationally: a smooth initial plateau followed by a gradual decline, without requiring a finite critical value of interaction density. In this sense, the present analysis supplies the mathematical mechanism that was missing from the original simulation study.

The large-population mean-field description is not, however, the complete story. Finite populations relax systematically more slowly during the principal transition region. Extensive simulations reveal that their trajectories can be approximately represented as temporally deformed versions of the mean-field evolution,

$$C_N(\tau) \approx C_\infty\left(\phi_N(\tau)\right),$$

with an effective clock satisfying $\phi_N'(\tau) < 1$ over substantial portions of the transient regime. This slowdown cannot be explained primarily by the elementary $O(1/N)$ correction arising from sampling agents without replacement. Instead, it is generated by finite-size distortions in the shape of the strategy distribution. At equal mean cooperation, finite populations can possess different variance, skewness, and higher cumulants than the corresponding mean-field distribution, thereby altering the instantaneous evolutionary drift.

The resulting mechanism forms a hierarchy. The mean drift of cooperation depends strongly on the variance V; the evolution of V depends on the third central moment K_3; and the evolution of K_3 depends on the fourth cumulant κ_4. Numerically, the discrepancy in variance production between finite populations and the mean-field reference is overwhelmingly associated with the skewness term, whereas the discrepancy in skewness dynamics is in turn overwhelmingly associated with the fourth cumulant. This explains why low-order moment closures can reproduce instantaneous derivatives reasonably well while failing when integrated over long intervals: the underlying process is genuinely distributional.

Rather than extending this hierarchy indefinitely, we therefore construct a reduced representation of the evolving distribution itself. Proper orthogonal decomposition of the mean-field distribution trajectory shows that it is highly compressible in Hellinger geometry. A small number of dominant modes captures most of its variation, and a four-dimensional quadratic latent model predicts later mean-field dynamics outside its fitting interval with low error. Finite-population distributions project close to the same reduced trajectory, but they traverse it more slowly and retain a smaller transverse deformation. This motivates the reduced representation

$$\dot{z_N} \approx v_N G_\infty\left(z_N\right) + r_N^\perp,$$

where z_N denotes the reduced distributional state, G_∞ the mean-field latent flow, v_N an effective finite-population evolutionary speed, and $r_N^\perp$ a transverse finite-size correction.

Crucially, the central geometric predictions of this framework were subsequently evaluated prospectively rather than only retrospectively. A frozen reduced representation and prespecified criteria were tested using 64 new simulations, 16 independent seeds for each of four population sizes, corresponding to approximately 1.404×10^9 additional pairwise encounters. Across all four populations, the direction of finite-population motion remained strongly aligned with the mean-field latent flow, with median directional cosines between approximately 0.984 and 0.994 and simultaneous intervals above the prespecified threshold of 0.95. Independently estimated effective clock speeds also agreed within the prespecified tolerance. A stronger prespecified requirement — that the transverse finite-size component decrease monotonically between every adjacent tested population size — remained statistically unresolved. Accordingly, the full composite hypothesis was not confirmed, although its two central dynamical predictions received prospective support.

2. Literature review

The study of cooperation spans evolutionary game theory, behavioral economics, biology, statistical physics, and complex-systems science because it addresses the tension between individually advantageous behavior and collectively efficient outcomes. Evolutionary game theory recast this problem as a population-dynamical one, asking how strategies change through selection, imitation, mutation, and stochastic drift [2–4]. Classical explanations of cooperation emphasize direct reciprocity, indirect reciprocity, kin selection, spatial clustering, network reciprocity, assortment, and other forms of population structure [5–9]. These mechanisms are important, but many rely on repeated relationships, reputation, spatial organization, or non-random matching. The model studied here deliberately removes those supports and retains anonymous random pairwise encounters, allowing encounter intensity and evolutionary updating to be examined in isolation.

That distinction motivated the preceding computational study by De los Reyes-Marín et al. [1], which examined up to 100,000 agents with continuous cooperation propensities, random matching, Fermi imitation, and mutation. The simulations produced a nonlinear cooperation landscape summarized empirically as C = F(I, μ): interaction density was the dominant predictor of observed cooperation, while mutation produced much smaller variation. The study also acknowledged that the relationship remained primarily simulation-based and that the analytical mechanism underlying the apparent regime changes had not yet been derived. The present paper addresses that unresolved question by asking whether I changes the evolutionary regime itself or instead changes the amount of evolutionary exposure accumulated before observation.

The relevant stochastic-process literature begins with pairwise comparison dynamics. Traulsen et al. [10] established Fermi updating as a finite-population evolutionary mechanism in which payoff differences and selection intensity determine probabilistic imitation. The current contribution therefore does not lie in introducing Fermi imitation, but in deriving the consequences of the specific sequential process used in De los Reyes-Marín et al. [1]. A second relevant literature concerns continuous strategy spaces. Ruijgrok and Ruijgrok [11] showed that evolutionary games with continuously distributed strategies can be expressed through distributional replicator-mutation dynamics. This provides a conceptual precedent for treating the cooperation-propensity distribution, rather than only its mean, as the population state.

Replicator theory also provides precedents for analytical reduction. Karev [12] showed that broad classes of replicator equations admit reduced descriptions and exponential tilting of distributions. Accordingly, the exponential solution obtained below is not presented as a new mathematical technique; its significance is that it emerges from the audited Fermi kernel and explains the interaction-density phenomenon previously observed computationally. Similarly, finite-size corrections and moment closure have an established history. Traulsen et al. [10] derived drift–diffusion descriptions for finite stochastic evolutionary processes, and Vardanyan and Saakian [13] used moment closure to study finite-population corrections to replicator dynamics. The present analysis differs by showing that, in the audited continuous-propensity process, the dominant finite-size discrepancy is not the direct pair-sampling correction but a distortion of the evolving strategy distribution.

This distributional effect naturally produces a moment hierarchy: the drift of mean cooperation depends on variance, variance production depends on the third central moment, and third-moment dynamics depend on the fourth cumulant. Moment hierarchies are familiar in stochastic population processes and statistical mechanics; the relevant contribution here is their explicit derivation from the microscopic Fermi rule and their quantitative connection to finite-population slowdown. Because low-order closures fail when integrated over time, the analysis subsequently moves from moment-space reduction to direct reduction of the probability distribution.

That step connects the study with reduced-order modeling. Proper orthogonal decomposition and related techniques are established methods for representing high-dimensional nonlinear dynamics in low-dimensional coordinates, and contemporary work has extended these ideas to probabilistic reduced-order systems [14]. Here these tools are used not as methodological novelties, but to test whether the apparently unbounded moment hierarchy evolves near a compact distributional trajectory. This is consistent with the broader complex-systems view that macroscopic regularities should be explained by explicit micro-to-macro mechanisms rather than by aggregate correlations alone [15–17].

The literature therefore supplies precedents for each individual ingredient—Fermi imitation, continuous-strategy distribution equations, exponential replicator solutions, finite-size analysis, moment closure, symbolic equation discovery, and reduced-order modeling [18]. What remains unresolved is their integration for the sequential anonymous Fermi process that generated the interaction-density landscape of De los Reyes-Marín et al. [1]. The gap addressed here is specific: to derive what interaction density controls, determine why the observed cooperation surface changes with I, and connect the microscopic process to a distributional mean-field flow and its finite-population corrections.

3. Methodology

The present study was designed as a mathematical and computational reconstruction of the evolutionary process underlying the cooperation landscape previously reported by De los Reyes-Marín et al. [1]. The objective was not to introduce a new behavioral mechanism or to modify the published model, but to recover the dynamical law implicit in its revised Fermi implementation and determine whether the empirical dependence $C = F(I, \mu)$ could be derived from the microscopic update rules. Accordingly, the analysis proceeded from the operational simulation code rather than from an exogenously imposed macroscopic equation. This distinction was necessary because the published narrative and the revised executable implementation are not fully aligned with respect to mutation timing: the model section describes mutation as conditional on low payoffs, whereas the revised results and computational implementation use pairwise Fermi imitation with mutation applied when imitation occurs. The present study therefore treats the revised Fermi code as the operative specification of the process and preserves this discrepancy explicitly as part of the audit trail.

The population consists of N agents, each characterized by a continuous cooperation propensity

$$p_i \in [0, 1].$$

During each microscopic encounter, two distinct agents i and j are sampled uniformly from the population. Their realized actions are generated independently according to

$$a_i \sim \text{Bernoulli}\left(p_i\right), \; a_j \sim \text{Bernoulli}\left(p_j\right).$$

Payoffs are assigned according to the Prisoner's Dilemma matrix. Unless otherwise stated, the baseline specification is

$$R = 3, \; S = 0, \; T = 5, \; P = 1,$$

satisfying

$$T > R > P > S.$$

These values correspond to the baseline environment used in the original computational study [1].

Evolutionary adaptation occurs immediately after each encounter. Let r_i and r_j denote the realized payoffs of the two agents. Agent i imitates agent j with Fermi probability

$$\Pr(i \leftarrow j) = \frac{1}{1 + \exp\left[-\beta\left(r_j - r_i\right)\right]},$$

while the reverse imitation probability is

$$\Pr(j \leftarrow i) = \frac{1}{1 + \exp\left[-\beta\left(r_i - r_j\right)\right]}.$$

The two imitation decisions are generated conditionally on the realized actions. When imitation occurs, the updating agent copies the partner's cooperation propensity as it existed before the encounter and adds a Gaussian perturbation,

$$p_i' = p_j + \mu Z, \; Z \sim \mathcal{N}(0, 1),$$

after which the value is clipped to $[0, 1]$. Agents that do not imitate retain their previous cooperation propensity. Updates are sequential: the modified population resulting from one encounter becomes the input state of the next encounter. There is no memory, reputation, persistent partner identity, punishment mechanism, or stable interaction network. A nominal generation is therefore only a block of I sequential encounters.

This protocol differs from accumulated-payoff updating, in which agents first play k encounters and then imitate using their total or average payoff. With summed payoffs, a deterministic k-fold increase in the payoff difference is equivalent in the Fermi rule to replacing β by $k\beta$; with random or averaged payoffs, the relationship also depends on payoff variability and normalization. Changing k can therefore change the transition kernel, fixation probabilities and stationary behavior. Here, I only groups already completed elementary updates, so it leaves that kernel unchanged.

The first methodological stage consisted of an exact reduction of the microscopic pairwise comparison process. Writing $x = p_i$, $y = p_j$, and

$$a = \tanh\left[\frac{\beta(T - S)}{2}\right],$$

averaging the Fermi imitation probability over the four possible joint action realizations yields

$$q(x, y) = \Pr(i \leftarrow j | x, y) = \frac{1}{2} + \frac{a}{2}(x - y).$$

For clarity, let $d = \beta(T - S)$ and $\ell(d) = \frac{1}{1+\exp(-d)}$. The four action outcomes CC, CD, DC and DD contribute, respectively,

$$q(x, y) = \frac{xy}{2} + x(1 - y)\ell(d) + (1 - x)y\ell(-d) + \frac{(1 - x)(1 - y)}{2},$$
$$\ell(\pm d) = \frac{1 \pm a}{2}, a = \tanh\left(\frac{d}{2}\right).$$

For CC and DD, both players receive the same payoff, so the payoff difference is zero and the imitation probability is 1/2. Substitution gives the reduced expression above and makes the cancellation of R and P explicit.

This reduction was derived algebraically rather than estimated from simulation data. It implies that, for the audited rule, R and P cancel from the marginal selection mechanism and that the payoff dependence enters through the composite quantity $\beta(T-S)$. Numerical controls using identical random sequences and payoff matrices preserving $T-S$ were used to verify the resulting process equivalence.

The role of interaction density was then studied at the transition-operator level. Let P denote the Markov transition operator corresponding to one pairwise encounter. Since I does not appear inside the one-encounter rule, the transition operator for one nominal generation is exactly

$$P_I = P^I.$$

After g generations,

$$\mathcal{L}\left(X_g\right) = \mathcal{L}\left(X_0\right) P^{gI}.$$

Thus, the process depends on the product $M = gI$, the cumulative number of encounters, rather than on the arbitrary partition of those encounters into generations. The natural time variable was consequently defined as expected participations per agent,

$$\boxed{\tau = \frac{2M}{N} = \frac{2gI}{N}.}$$

This relation was treated as an exact clock identity, not as a fitted scaling law. Coupled simulations using the same random sequence but different groupings of a fixed number of encounters were used as a computational verification: at equal M, final population states coincided trajectory by trajectory.

For positive mutation, the stationary finite-population process was analyzed through a minorization argument. Because payoffs and β are finite, imitation and non-imitation probabilities are uniformly bounded away from zero. Gaussian mutation with clipping retains a strictly positive interior density over $(0,1)$, and a finite sequence of ordered encounters can update all coordinates with positive probability. These properties yield a Doeblin-type minorization for an appropriate power of the transition kernel and imply a unique aperiodic stationary distribution for the finite process with $\mu > 0$. Since any invariant distribution of P is also invariant under P^I, the stationary cooperation level is common to every $I > 0$,

$$C^*(I; N; \mu; \kappa) = C^*(N, \mu, \kappa),$$

and therefore

$$\frac{\partial C^*}{\partial I} = 0.$$

This result concerns the stationary process and does not imply that finite simulation horizons are sufficient to reach stationarity.

The next stage derived the exact conditional drift of mean cooperation. Defining

$$C = \frac{1}{N}\sum_{i=1}^{N} p_i, V = \frac{1}{N}\sum_{i=1}^{N}\left(p_i - C\right)^2,$$

and

$$b_\mu(p) = \mathrm{E}\left[\mathrm{clip}\left(p + \mu Z, 0, 1\right)\right] - p,$$

the finite-N mutation quantities were defined as

$$B = \overline{b_\mu(p)},\ D = \overline{(p-C)\, b_\mu(p)},\ k_N = \frac{N}{N-1}.$$

The exact one-encounter conditional drift, expressed per unit of interaction time $2/N$, is

$$\boxed{\frac{\mathbb{E}\left[\Delta C \mid \boldsymbol{p}\right]}{2/N} = -a k_N V + \frac{1}{2}B - \frac{a}{2}k_N D.}$$

This identity was evaluated directly on saved population states and independently checked through finite-interval integration. Its central implication is that mean cooperation alone is not a sufficient state variable: two populations with identical C but different internal variance or mutation-related covariances have different expected drifts.

The variance dependence has a direct selection interpretation. In the mutation-free mean-field equation, define the effective selection rate $w(p) = -ap$. The resulting covariance identity is

$$\dot{C} = \mathrm{Cov}_f(p, w) = -aV.$$

This is the selection-by-covariance structure of Price [19]. For constant a, the mean effective rate satisfies $\frac{d\mathrm{E}[w]}{d\tau} = \mathrm{Var}(w) = a^2 V$, connecting it to the variance contribution emphasized by Fisher's fundamental theorem [20]. The effective rate w is not the material game payoff. With mutation and clipping, the additional B and D terms in the exact drift above must be retained.

The large-population limit was formulated as a distributional evolution equation. Let f_τ denote the probability distribution of cooperation propensities on $[0,1]$, let

$$C_\tau = \int_0^1 p\, f_\tau(dp),$$

and let $K_\mu(du \mid p)$ denote the exact distribution of

$$\mathrm{clip}(p + \mu Z, 0, 1),$$

including the boundary atoms created by clipping. Under the mean-field approximation, the evolution equation is

$$\boxed{\partial_\tau f_\tau(du) = \int_0^1 \left[\frac{1}{2} + \frac{a}{2}\left(C_\tau - p\right)\right] K_\mu(du \mid p)\, f_\tau(dp) - \left[\frac{1}{2} + \frac{a}{2}\left(u - C_\tau\right)\right] f_\tau(du).}$$

Its first moment reproduces the large-N limit of the finite-population drift equation. No claim of uniform propagation of chaos for arbitrarily long times is made; the mean-field equation is treated as the deterministic large-population reference against which finite-N simulations are compared.

For $\mu = 0$, the mutation kernel reduces to $K_0(du \mid p) = \delta_p(du)$, allowing an analytical solution. For an arbitrary initial density,

$$f_\tau(p) = \frac{f_0(p)\, e^{-a\tau p}}{Z(\tau)},\quad Z(\tau) = \int_0^1 f_0(u)\, e^{-a\tau u}\, du.$$

For the uniform initialization used in the original simulations,

$$f_0(p) = 1,$$

the mean cooperation level becomes

$$\boxed{C(\tau) = \frac{1}{a\tau} - \frac{1}{e^{a\tau} - 1}.}$$

Near the origin,

$$C(\tau) = \frac{1}{2} - \frac{a\tau}{12} + \frac{(a\tau)^3}{720} + O\left(\tau^5\right),$$

whereas for large exposure,

$$C(\tau) \sim \frac{1}{a\tau}.$$

Substitution of $\tau = 2gI/N$ provides an explicit analytical relation between interaction density and cooperation under the mutation-free mean-field assumptions. Because the published observable was based on cooperation averaged across a final generational window rather than solely on the terminal state, the theoretical comparison also used the corresponding time-window average where applicable, avoiding comparisons between an instantaneous theoretical state and a temporally averaged simulation observable.

For $\mu > 0$, the distribution equation was solved numerically using a finite-volume representation of $[0,1]$ that treats the clipping masses at $p = 0$ and $p = 1$ as explicit boundary atoms and represents the interior by midpoint cells. The mutation matrix was constructed from exact Gaussian probabilities integrated over the finite-volume cells. Time integration used an adaptive BDF solver on the probability simplex, with the final mass reconstructed from the unit-mass constraint rather than by post hoc renormalization. Numerical validity was assessed through mass conservation, non-negativity, residual norms, spatial refinement, and independent temporal-tolerance refinement.

The most stringent numerical certification was performed for the case $\mu = 0.02$, $\beta = 1$, with all agents initially cooperating. Spatial refinement from 1281 to 2561 nodes satisfied the pre-specified convergence criteria, and an independent strict-to-ultra temporal comparison at 2561 nodes also passed. The maximum temporal-sensitivity difference in mean cooperation was approximately 2.11×10^{-4}, while the final ultra-strict L_1 residual was approximately 2.17×10^{-8}. These checks establish numerical sensitivity control for the discretized mean-field solution under the chosen tolerances; they are not interpreted as rigorous continuum-error bounds.

Finite-population relaxation was then investigated directly in interaction time τ, rather than in nominal generations. Independent stochastic simulations were performed over multiple population sizes, with trajectories recorded at common values of τ. Cooperation, variance, higher moments, boundary masses, and complete histograms of the cooperation-propensity distribution were stored at each checkpoint. First-passage times through cooperation levels $C = 0.75, 0.50, 0.25$, and 0.10 were estimated by interpolation between consecutive checkpoints and compared with the corresponding mean-field crossing times. In parallel, fixed-τ deviations

$$E_N(\tau) = \mathbb{E}\left[C_N(\tau)\right] - C_\infty(\tau)$$

were analyzed to distinguish genuine state deviations from amplification caused specifically by first-passage observables.

The finite-size mechanism was studied analytically through a hierarchy of moment identities. Exact conditional evolution equations were derived for the first and second raw moments and for the finite-population variance, including the quadratic-variation contribution generated by stochastic encounter-level updates. In mean-field form, the variance equation is

$$\frac{dV}{d\tau} = -aK_3 + \frac{1}{2}B_2 - CB_1 - \frac{a}{2}D_2 + aCD_1,$$

where K_3 is the third central moment and B_1, B_2, D_1, D_2 collect exact mutation and clipping contributions. The analogous equation for the third central moment is

$$\frac{dK_3}{d\tau} = -a\left(K_4 - 3V^2\right) + \frac{1}{2}\left[B_3 - 3CB_2 + 3\left(C^2 - V\right)B_1\right] - \frac{a}{2}\left[D_3 - 3CD_2 + 3\left(C^2 - V\right)D_1\right].$$

Thus, the fourth cumulant

$$\kappa_4 = K_4 - 3V^2$$

appears explicitly as the next term in the hierarchy. These equations were used to decompose finite-size drift discrepancies at equal C into direct finite-N corrections and differences produced by the shape of the population distribution.

Several closure approximations were examined diagnostically. A small-mutation interior approximation was obtained by neglecting boundary-clipping contributions, giving

$$\dot{C} \approx -aV,$$

$$\dot{V} \approx -aK_3 + \frac{\mu^2}{2},$$

and

$$\dot{K}_3 \approx -a\kappa_4 - \frac{3}{2}a\mu^2 V.$$

Gaussian, Beta-based, Pearson-type, and empirical closures were compared against the exact instantaneous drift and then integrated as reduced systems. Their performance was evaluated separately at the derivative level and at the trajectory level because a closure that approximates an instantaneous derivative accurately can still accumulate substantial dynamical error when integrated. The failure of stable low-order moment closure motivated a transition from moment-space reduction to direct distributional reduction. These interior approximations require negligible probability near the clipping boundaries and are not uniform in time, especially from the all-cooperating initial state. They are diagnostic closures only; the full mean-field equation and its numerical reference retain the boundary atoms and exact clipping terms.

Reduced-order distribution dynamics were constructed from the certified mean-field probability distributions using proper orthogonal decomposition. Distributions were represented on 128 bins, and the POD decomposition was performed on square-root probabilities,

$$h_k = \sqrt{f_k},$$

so that Euclidean distances in the transformed space are directly related to Hellinger geometry. The distribution was approximated as

$$h(\tau) \approx \overline{h} + \sum_{j=1}^{m} z_j(\tau)\psi_j.$$

Here f_k is the probability mass of bin k, not a density value. In the 128-bin POD input, the atom at p = 0 is included in the first bin and the atom at p = 1 in the last bin, together with the corresponding interior cell masses; they are not separate POD coordinates. The finite-volume reference keeps the atoms separate before this mass-preserving aggregation. Square-root coordinates therefore define a valid discrete Hellinger geometry, but do not resolve an endpoint atom from nearby interior mass within its terminal bin. This same aggregation and the frozen POD basis were used for the finite-population comparisons.

The number of modes was not selected solely from cumulative explained variance. Temporal block cross-validation was used to assess reconstruction error in Hellinger distance, Wasserstein distance, mean cooperation, variance, and skewness. Although four modes accounted for approximately 98.6% of the mean-field trajectory energy, higher-dimensional reconstructions were required for high-precision recovery of the full distribution. Four modes were retained for the latent dynamical model because they provided a compact state representation while preserving the dominant trajectory geometry.

A quadratic latent vector field

$$\dot{z} = G(z)$$

was estimated from the four POD coordinates. Ridge regularization was selected by temporal cross-validation of the latent derivatives. The model was trained on an early portion of the mean-field trajectory and then integrated beyond the fitting window, so that the late-time comparison constituted a temporal prediction rather than an in-sample reconstruction. The resulting distributions were reconstructed from latent coordinates, squared to recover non-negative probability masses, and renormalized to unit mass. Predictive performance was evaluated using Hellinger distance, Wasserstein distance, and errors in C, V, and K_3.

To compare finite populations with the reduced mean-field trajectory, each finite-N histogram was matched to the mean-field distribution minimizing Hellinger distance. This defines a distributional phase

$$\phi_N(\tau) = \arg\min_s d_H\left[f_N(\tau), f_\infty(s)\right].$$

A monotonicity constraint was imposed on the estimated phase because evolutionary phase was interpreted as cumulative progress along the reference trajectory. The resulting lag was

$$\delta_N(\tau) = \tau - \phi_N(\tau).$$

The reduction in Hellinger and Wasserstein distances before and after phase matching quantified the tangential part of the finite-size discrepancy, whereas the remaining distance quantified deformation transverse to the mean-field trajectory. New finite-N simulations were projected onto the frozen four-mode basis, allowing direct comparison between finite-population motion and the mean-field latent tangent.

The local finite-population latent velocity was estimated from smoothed derivatives of the POD coordinates. Let $z_N(\tau)$ denote the finite-N latent state and G_∞ the tangent to the mean-field trajectory at the matched phase. The tangential speed was estimated as

$$\alpha_N = \frac{\langle z_N', G_\infty \rangle}{\|G_\infty\|^2},$$

while directional alignment was quantified through

$$\cos\theta_N = \frac{\langle z_N', G_\infty \rangle}{\|z_N'\| \, \|G_\infty\|}.$$

The non-tangential component was measured by

$$r_N^\perp = \dot{z}_N - \alpha_N G_\infty,$$

and its relative magnitude by

$$\frac{\| r_N^\perp \|}{\| \dot{z}_N \|}.$$

The independently estimated distributional clock speed

$$v_N = \frac{d\phi_N}{d\tau}$$

was then compared with α_N. Agreement between these quantities was interpreted as geometric consistency between the phase-based and flow-based measurements of finite-population evolutionary speed, rather than as two statistically independent causal tests.

The final stage was a prospective validation using frozen analytical and reduced-order components. Before the confirmatory simulations were launched, the tested population sizes, random seeds, analysis pipeline, thresholds, and decision rules were written to a prospective protocol file and sealed with a SHA-256 hash. This served as an internal reproducibility control rather than an externally registered preregistration. Four population sizes were tested,

$$N = 25000, 75000, 250000, 1000000,$$

with 16 previously unused random seeds for each N, yielding 64 new simulations and approximately

$$1.404 \times 10^9$$

additional pairwise encounters. The latent POD basis, quadratic flow model, phase estimator, and primary analysis rules were frozen before these simulations were analyzed.

Three families of prospective criteria were specified. First, the median directional cosine between finite-N latent motion and the matched mean-field tangent was required to exceed

$$0.95$$

for each tested population size. Second, the absolute discrepancy between the tangential-speed estimate α_N and the independently estimated phase speed $d\phi_N/d\tau$ was required to remain below

$$0.10.$$

Third, the transverse finite-size component was required to decrease between each pair of consecutively larger populations. Uncertainty was assessed by resampling entire independent seeds rather than individual trajectory observations. Five thousand bootstrap replications were used, with Bonferroni adjustment across the 11 primary confirmatory contrasts. A criterion was classified as supported only when the corresponding simultaneous confidence interval lay entirely within the respecified acceptance region; intervals crossing the threshold were classified as indeterminate rather than as evidence of equivalence or contradiction.

All stochastic experiments used explicit independent seeds, and outputs included raw trajectory data, probability histograms, analytical diagnostics, numerical convergence checks, manifests, and SHA-256 integrity records. Completed runs were preserved independently to allow interrupted campaigns to resume without counting previously completed seeds as new evidence. Exploratory and confirmatory analyses were kept separate: model development, moment-hierarchy analysis, reduced-order construction, and finite-size mechanism discovery preceded the frozen prospective test, whereas the final confirmatory campaign did not refit the reduced representation or modify the prespecified thresholds in response to its outcomes.

The methodological strategy can therefore be summarized as a progression from microscopic rules to macroscopic dynamics rather than as a sequence of independent simulation exercises. The one-encounter stochastic process was first reduced analytically; interaction density was identified at the transition-operator level; exact conditional drift identities were derived; the corresponding mean-field distribution equation was solved analytically where possible and numerically where necessary; finite-size discrepancies

were decomposed through moment and distributional diagnostics; the resulting distributional dynamics were compressed into a reduced latent representation; and the principal geometric predictions of that representation were finally subjected to prospectively generated simulations with frozen criteria. This design was intended to transform the empirical relationship reported by De los Reyes-Marín et al. [1] into a testable dynamical theory rather than merely fit a more elaborate functional form to the original cooperation landscape.

4. Results

The analysis first established that the apparent interaction-density effect reported in the preceding computational study can be reduced to an exact property of the microscopic transition process. In the audited sequential Fermi implementation, interaction density I does not modify the one-encounter transition rule; it only determines how many encounters are executed within a nominal generation. If P denotes the transition operator associated with one encounter, then a generation containing I encounters is governed by

$$P_I = P^I,$$

and after g generations,

$$\mathcal{L}\left(X_g\right) = \mathcal{L}\left(X_0\right) P^{gI}.$$

Consequently, the relevant time variable is the accumulated interaction exposure

$$\boxed{\tau = \frac{2gI}{N},}$$

rather than the generation index alone. Numerical controls in which the same total number of encounters was partitioned differently across generations produced identical population states under the same random stream. Thus, changing I while holding g fixed changes the evolutionary time at which the population is observed. For positive mutation, the finite stochastic process possesses a unique stationary distribution shared by every $I > 0$, implying

$$\boxed{C^*(I; N; \mu; \kappa) = C^*(N, \mu, \kappa)}$$

and therefore

$$\boxed{\frac{\partial C^*}{\partial I} = 0.}$$

The decline in cooperation previously associated with increasing interaction density must therefore be interpreted as a finite-horizon relaxation effect within the audited model rather than as evidence of an I-induced stationary bifurcation.

The microscopic Fermi process also admitted an exact reduction of the payoff dependence. Averaging over the possible action realizations gives

$$q(x, y) = \frac{1}{2} + \frac{a}{2}(x - y),$$

with

$$\boxed{a = \tanh\left[\frac{\beta(T - S)}{2}\right].}$$

The reward R and punishment P cancel from the marginal pairwise selection mechanism, so matrices preserving $\beta(T - S)$ generate the same selective kernel. This result substantially reduces the effective parameter space and clarifies why several payoff scenarios in the original simulations produced very similar macroscopic cooperation trajectories.

An exact finite-population conditional drift equation was then obtained for mean cooperation. Defining C as the mean cooperation propensity, V as its population variance, B as the mean clipping-induced mutation bias, D as its covariance with the centered cooperation propensity, and $k_N = N/(N - 1)$, the drift per unit interaction time is

$$\boxed{\frac{\mathbb{E}[\Delta C \mid \boldsymbol{p}]}{2/N} = -ak_N V + \frac{1}{2}B - \frac{a}{2}k_N D.}$$

This identity demonstrates that C alone does not close the evolutionary dynamics. Populations with identical mean cooperation but different internal distributions have different expected drifts. The empirical relationship $C = F(I, \mu)$ observed in the original study is therefore a reduced macroscopic projection of a higher-dimensional distributional process rather than the fundamental dynamical law itself.

In the mutation-free mean-field limit, this distributional process admits an explicit analytical solution. For an initially uniform population,

$$f_\tau(p) = \frac{e^{-a\tau p}}{Z(\tau)},$$

and the mean cooperation level is

$$\boxed{C(\tau) = \frac{1}{a\tau} - \frac{1}{e^{a\tau} - 1}.}$$

Near $\tau = 0$,

$$C(\tau) = \frac{1}{2} - \frac{a\tau}{12} + \frac{(a\tau)^3}{720} + O\left(\tau^5\right),$$

while for large interaction exposure,

$$C(\tau) \sim \frac{1}{a\tau}.$$

Substitution of $\tau = 2gI/N$ gives an explicit analytical relation between cooperation and interaction density,

$$\boxed{C(g, I) = \frac{N}{2agI} - \frac{1}{\exp(2agI/N) - 1},}$$

under the mutation-free mean-field assumptions. The analytical trajectory reproduces the qualitative structure observed in the original simulations: an initially flat region followed by a smooth decline as interaction exposure accumulates. No singular critical interaction density is required to generate this pattern.

For positive mutation, the mean-field distribution equation was solved numerically and subjected to increasingly strict convergence controls. In the representative case $\mu = 0.02$, $\beta = 1$, with all agents initially cooperating, spatial refinement from 1281 to 2561 nodes passed the predefined convergence criteria, while an independent strict-to-ultra temporal refinement on the 2561-node grid also passed. The maximum temporal-sensitivity difference in C was 2.11×10^{-4}, and the final ultra-strict L_1 residual was 2.17×10^{-8}. This provided a numerically certified reference trajectory for subsequent finite-population comparisons.

The finite-population simulations revealed a systematic transient delay relative to this mean-field reference. In the first-passage analysis, all replicates crossed each tested cooperation level, but the finite-N crossings occurred progressively later than the deterministic prediction. For example, at $N = 1000$, the mean delay relative to mean-field was approximately 51.34 interaction-time units for $C = 0.10$, 40.51 for $C = 0.25$, and 20.13 for $C = 0.50$. At $N = 500000$, the corresponding delays had fallen to approximately 13.92, 8.70, and 2.19, respectively. The delay therefore weakened substantially as population size increased, although the rate of convergence depended strongly on the location of the crossing within the transient.

The effective-clock analysis confirmed this interpretation. A nonparametric phase map was defined by

$$\phi_N(\tau) = C_\infty^{-1}\left[C_N(\tau)\right],$$

and its derivative was interpreted as an effective evolutionary speed,

$$v_N(\tau) = \frac{d\phi_N}{d\tau}.$$

Across the transition region, $v_N(\tau)$ was consistently below one. For example, at $N = 100000$, the estimated speeds were approximately 0.453, 0.638, and 0.879 at $\tau = 60, 90$, and 120, respectively, with bootstrap confidence intervals entirely below one. At $N = 500000$, the corresponding values were approximately 0.530, 0.805, and 0.887. The accumulated lag therefore increased during the main transition and diminished as the finite trajectory recovered the mean-field rate.

The mechanism underlying this slowdown was identified by comparing finite and mean-field drift terms at equal values of C. The direct $O(1/N)$ correction caused by sampling distinct pairs accounted for a median fraction of only 6.80×10^{-5} of the total drift discrepancy over the transition window. By contrast, the difference in the shape of the strategy distribution accounted for essentially the entire gap, with a median relative contribution of approximately 1.00007. The speed inferred independently from the exact drift ratio,

$$\frac{A_N}{A_{\mathrm{MF}}},$$

correlated at

$$\boxed{r = 0.989779}$$

with the effective-clock speed obtained from the phase analysis. The deficit in evolutionary speed also correlated at

$$\boxed{r = 0.987214}$$

with the deficit in variance relative to the mean-field distribution at equal cooperation.

The variance deficit was itself traced to higher-order distributional structure. The exact variance equation showed that its evolution depends on the third central moment K_3. At equal C, the component associated with K_3 explained a median of approximately 96.5% of the finite-versus-mean-field gap in variance production, whereas mutation terms contributed approximately 1.5% and covariance terms less than 0.4%. The correlation between the observed variance-drift gap and the K_3 component was

$$\boxed{r = 0.999863}.$$

The same structure continued one level higher. The mean-field equation for the third central moment contains the fourth cumulant

$$\kappa_4 = K_4 - 3V^2,$$

and the finite-size discrepancy in K_3 dynamics was overwhelmingly associated with this term. The fourth-cumulant component explained a median of approximately 97.5% of the total K_3-drift gap, with

$$\boxed{r = 0.999955}$$

between the total gap and the fourth-cumulant contribution. Direct finite-N chain corrections were negligible. The exploratory scatter plots supporting these correlations and the effective-speed diagnostics are provided in Fig. A.2. Their points are population-size/time summaries, rather than independent simulation replicates.

These results established a mechanistic hierarchy,

$$\boxed{\kappa_4 \to \dot{K}_3 \to K_3 \to \dot{V} \to V \to \dot{C},}$$

but also demonstrated that conventional low-order moment closures were insufficient for reproducing the complete integrated trajectory. Although Gaussian and Beta-type closures reproduced some instantaneous derivatives with moderate accuracy, integration of the resulting three-dimensional systems accumulated large errors outside the fitting region. A flexible empirical closure performed nearly perfectly in-sample but failed catastrophically in temporal extrapolation, confirming that local predictability of higher moments does not guarantee a stable reduced dynamical system.

The failure of low-order closure motivated a direct reduced-order representation of the probability distribution. Proper orthogonal decomposition in square-root-probability coordinates revealed a strongly concentrated spectrum. The first mode explained approximately 61.3% of the trajectory variation, two modes 86.2%, three modes 95.4%, four modes 98.6%, and six modes more than 99.9%. High-precision reconstruction of the full distribution required more modes than suggested by cumulative explained variance alone: temporal block cross-validation selected ten modes for the most accurate distributional reconstruction, reducing the cross-validated Wasserstein error to approximately 8.24×10^{-4}. Fig. A.1 displays the four retained spatial modes. The first primarily contrasts low and high propensities; the second contrasts the extremes with the interior and changes spread; the third and fourth capture additional asymmetric and near-boundary redistributions. These are mixed distributional deformations, not modes constrained to correspond one-to-one to individual moments.

Despite this requirement for precise reconstruction, a four-dimensional latent state proved sufficient to reproduce the dominant dynamics. A quadratic latent vector field trained on the earlier portion of the mean-field trajectory remained predictive after the training interval. In the out-of-sample temporal region, the latent model achieved an RMSE of approximately 0.0395 in Hellinger distance, 0.00406 in Wasserstein distance, 0.00403 in mean cooperation, 0.00128 in variance, and 8.73×10^{-4} in the third central moment. Unlike the integrated moment closures, the reduced distributional representation therefore retained useful predictive stability beyond its fitting window.

The distribution-level analysis independently confirmed the finite-population lag. Matching each finite-N histogram to the closest mean-field distribution in Hellinger distance produced median distributional lags of approximately 24.75, 21.50, 15.75, and 13.50 for $N = 2000$, 10000, 100000, and 500000, respectively. Phase alignment reduced the median Hellinger discrepancy by approximately 68%–74%. At $N = 500\,000$, for example, the median Hellinger distance fell from approximately 0.1468 at equal nominal τ to 0.0452 after phase alignment. The residual did not vanish, showing that finite population size produces both a dominant tangential delay and a smaller transverse deformation of the strategy distribution.

Projection of finite-population trajectories onto the frozen four-dimensional mean-field latent space provided a direct geometric test of this interpretation. Over the principal transition window, the median cosine between the finite-N velocity and the matched mean-field tangent was

$$0.9660,\ 0.9686,\ 0.9875,\ 0.9908$$

for $N = 2000, 10000, 100000$, and 500000, respectively. The corresponding transverse fractions of velocity were approximately $0.259, 0.249, 0.157$, and 0.135. The finite-population dynamics therefore became progressively more aligned with the mean-field trajectory as N increased. The speed obtained by projecting finite-N motion onto the mean-field tangent also remained close to the independently estimated distributional phase speed.

These exploratory results motivated a final prospective validation using frozen models, new random seeds, and prespecified acceptance criteria. The confirmatory campaign consisted of 64 new simulations across

$$N = 25000, 75000, 250000, 1000000,$$

with 16 independent seeds per population size and a total of

$$1.404 \times 10^9$$

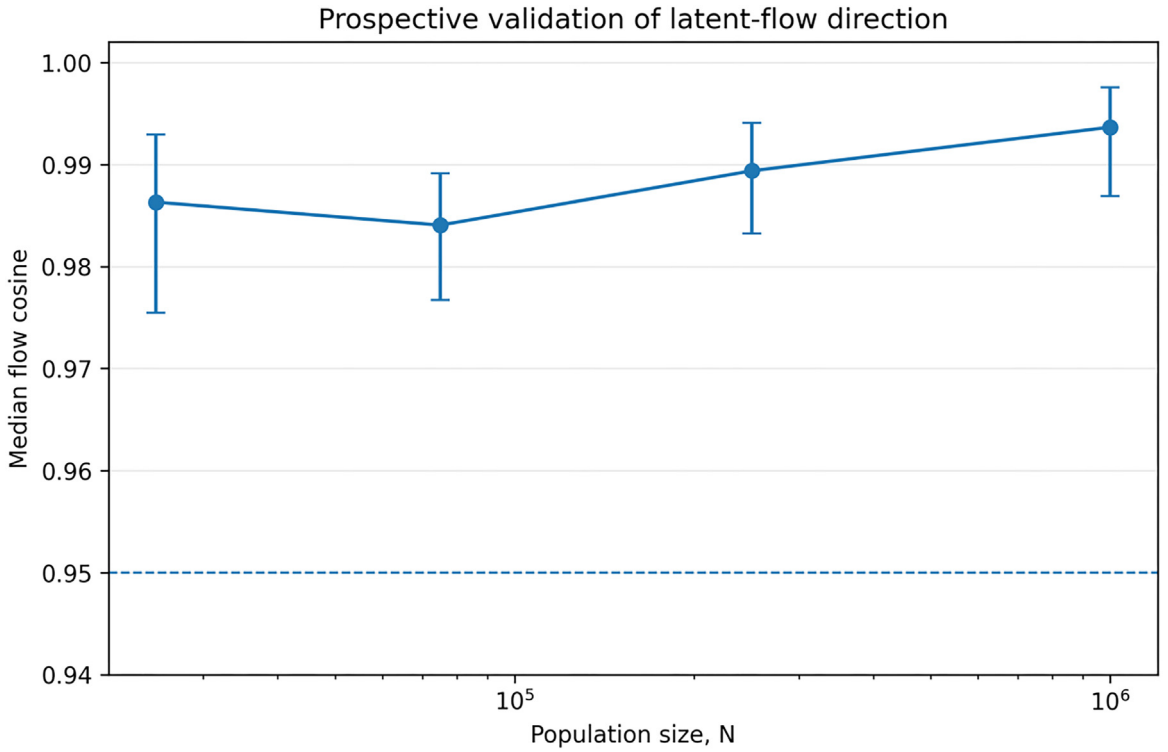


Fig. 1. Prospective validation of latent-flow direction. Median cosine between the finite-population latent velocity and the phase-matched mean-field tangent for the four prospectively tested population sizes. Error bars represent Bonferroni-adjusted simultaneous intervals obtained by bootstrap resampling of complete independent seeds. The dashed horizontal line indicates the prespecified acceptance threshold of 0.95; all four population sizes satisfy the confirmatory criterion.

new pairwise encounters. The reduced basis, latent flow, phase estimator, primary thresholds, and statistical decision rules were frozen before the simulations were analyzed.

The first confirmatory criterion required the median directional cosine between finite-population motion and the matched mean-field latent tangent to exceed 0.95. This prediction was supported at all four population sizes. The observed values were

$$0.9863,\ 0.9841,\ 0.9894,\ 0.9937,$$

and all Bonferroni-adjusted simultaneous intervals remained entirely above the prespecified threshold. The second criterion required the absolute difference between the tangent-derived speed and the independently estimated phase speed to remain below 0.10. This criterion was also supported at all four population sizes, with estimates of approximately

$$0.0495,\ 0.0663,\ 0.0620,\ 0.0608,$$

and simultaneous intervals entirely below the prespecified tolerance. Fig. 2 shows the agreement between the two independently constructed effective-speed estimates. Fig. 1 summarizes the prospectively validated directional alignment.

The third confirmatory criterion required the transverse velocity fraction to decrease strictly between every pair of consecutively larger populations. The observed median transverse fractions were

$$0.1649,\ 0.1779,\ 0.1454,\ 0.1125,$$

for $N = 25000, 75000, 250000,$ and 1000000, respectively. Although the largest population exhibited the smallest transverse component and the overall pattern remained compatible with decreasing finite-size distortion, none of the three pairwise decrease contrasts achieved a simultaneous interval lying entirely above zero. All three were therefore classified as indeterminate. Fig. 3 displays the non-monotonic but overall declining pattern of the transverse component.

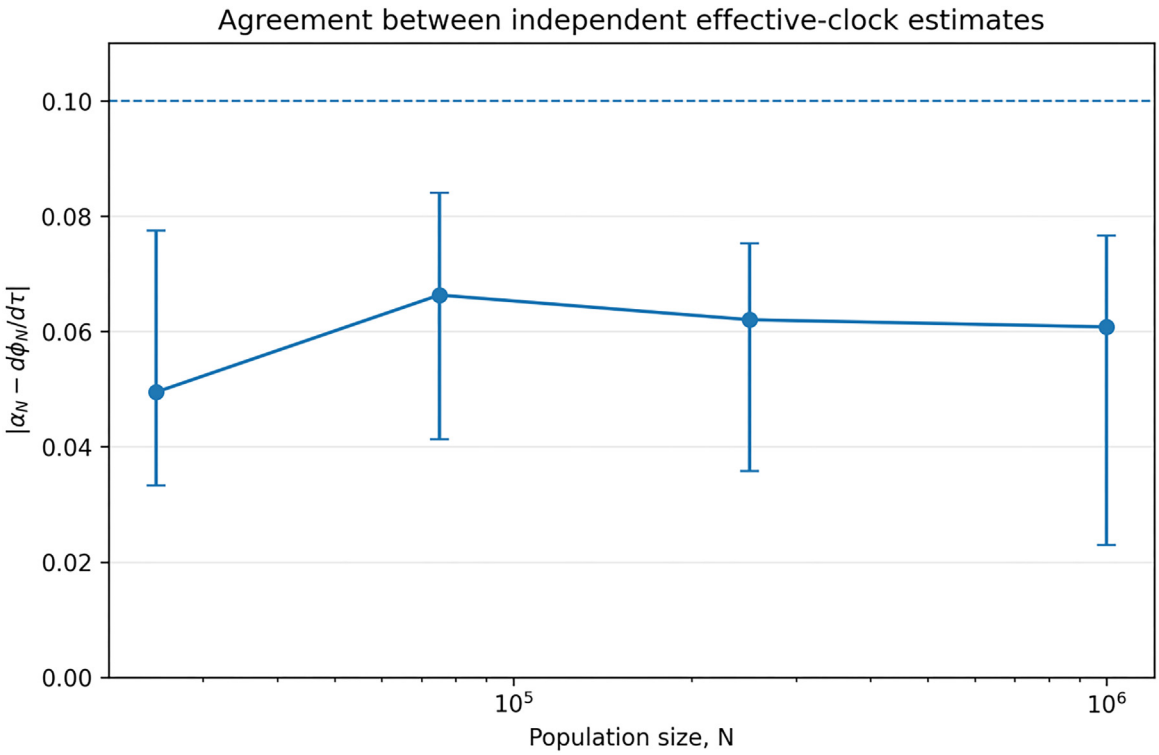


Fig. 2. Agreement between independent estimates of effective evolutionary speed. Absolute difference between the tangential speed alpha_N, obtained from projection of finite-population latent motion onto the mean-field tangent, and the independently estimated distributional phase speed dphi_N/dtau. Error bars represent Bonferroni-adjusted simultaneous intervals obtained from bootstrap resampling of complete independent seeds. The dashed horizontal line indicates the prespecified tolerance of 0.10; all four tested population sizes satisfy the confirmatory criterion.

Accordingly, the prespecified composite hypothesis was **not fully confirmed**, because one of its three required components remained statistically unresolved. It was not rejected either. The two central dynamical predictions — the common direction of reduced flow and agreement between independent measures of evolutionary speed — received prospective support across all tested population sizes, whereas strict pairwise monotonic decay of the transverse correction remained undetermined.

5. Discussion

The central result of this study is a reinterpretation of the interaction-density effect reported by De los Reyes-Marín et al. [1]. The earlier simulations correctly identified a strong nonlinear dependence of observed cooperation on I, but the present analysis shows that, in the audited sequential Fermi implementation, I does not modify the one-encounter transition kernel. It determines how often that kernel is applied before observation. The exact clock $\tau = 2gI/N$ therefore separates structural change from temporal rescaling: at fixed generation count, larger I means later evolutionary exposure. For positive mutation, all $I > 0$ share the same finite-population stationary distribution, so the decline observed at high I is better interpreted as finite-horizon relaxation than as an I-induced stationary bifurcation.

This interpretation also clarifies the original machine-learning result. Interaction density accounted for almost 90% of Random Forest feature importance in the preceding study. A parameter that changes the number of microscopic update events by orders of magnitude will naturally become highly predictive of the state observed at a fixed nominal horizon. Feature importance therefore captured a genuine statistical dependency, but it could not distinguish a structural parameter effect from a hidden change in effective time. The mutation-free solution makes the distinction explicit: the same plateau-and-decline pattern arises from smooth relaxation along τ without a singular critical value of I.

The exact Fermi reduction further shows that the selective kernel depends on β(T-S), while R and P cancel from the marginal pairwise imitation probability. This helps explain why several payoff scenarios in the earlier simulations had similar macroscopic trajectories, but the conclusion is specific to the audited update rule and should not be generalized to other evolutionary protocols.

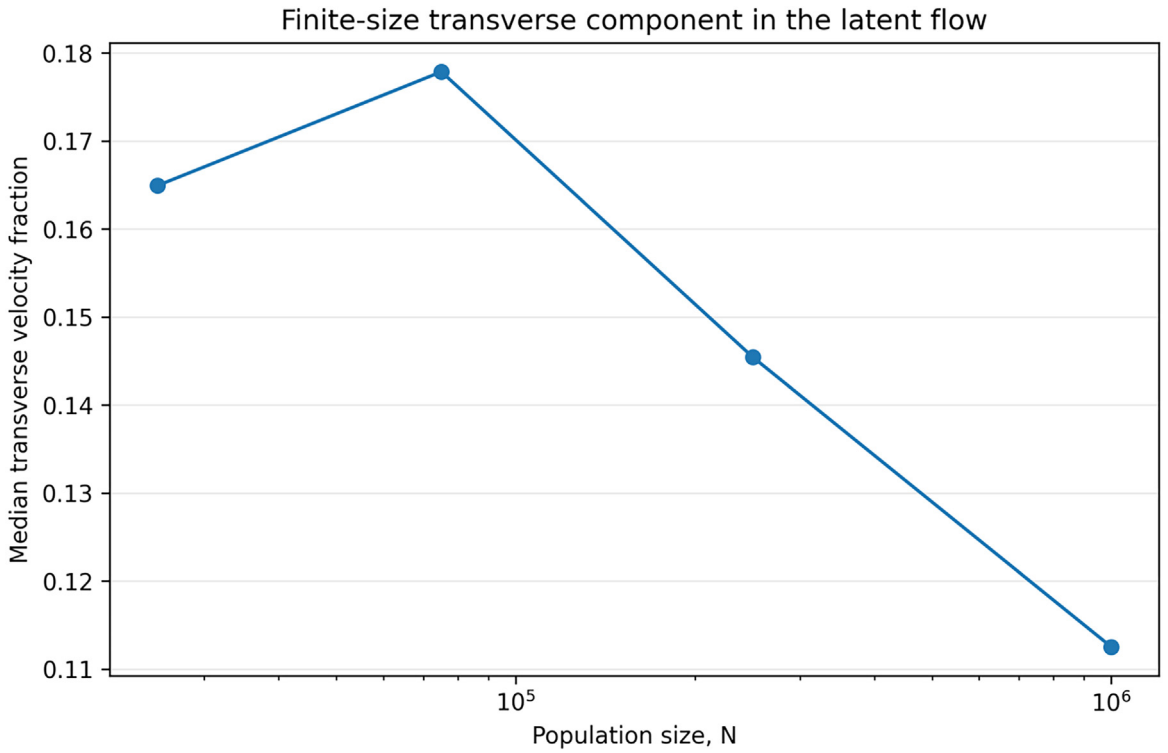


Fig. 3. Finite-size transverse component of the latent evolutionary flow. Median fraction of finite-population latent velocity orthogonal to the phase-matched mean-field tangent across the four prospectively tested population sizes. The largest population exhibits the smallest transverse component, but the sequence is not strictly monotonic. Bonferroni-adjusted pairwise contrasts did not establish monotonic decline, so this component of the prespecified composite hypothesis remained indeterminate.

More importantly, the exact drift equation establishes that C alone is not a sufficient state variable. Populations with the same mean cooperation can have different expected drifts because their variance and mutation-related distributional terms differ. The correct macroscopic object is therefore the evolving distribution f(p, τ), not the mean in isolation.

Finite populations depart from the deterministic mean-field trajectory in a strongly time-dependent way. The direct O(1/N) pair-sampling correction explains only a negligible fraction of the drift gap; almost all of the discrepancy is associated with differences in distributional shape. The moment analysis identifies the mechanism: the third central moment tracks the gap in variance production with r = 0.999863, and the fourth cumulant tracks the gap in third-moment dynamics with r = 0.999955. These correlations are not universal causal laws, but they show that the slowdown is propagated through higher-order structure rather than through a simple correction to the mean. The failure of stable two- and three-moment closures reinforces the conclusion that the process is genuinely distributional. Stochastic imitation can also remove strategy lineages and deplete diversity, the analogue of genetic drift and coalescence in this copying process. Algebraically, $V = M_2 - C^2$ makes the exact variance increment contain the non-positive term $-\mathrm{E}\left[(\Delta C)^2 \mid p\right]$, in addition to the changes in the raw moments. Its accumulated distributional effects need not be captured by the small instantaneous pair-sampling correction. This explains why a variance deficit can slow selection, without treating the observed correlations as a complete genealogical or asymptotic theory.

The reduced-order analysis provides a useful alternative. The mean-field trajectory is highly compressible, although not exactly four-dimensional: four POD modes explain about 98.6% of its variation, whereas more modes are needed for high-precision distribution reconstruction. A four-dimensional quadratic latent model nevertheless predicts later mean-field dynamics outside its fitting window. At the finite-population level, phase matching removes roughly 68%–74% of the Hellinger discrepancy, indicating that most of the finite-size error is tangential — a delay along a similar distributional path — while a smaller transverse deformation remains.

This interpretation was tested prospectively rather than only retrospectively. In the prospectively frozen confirmatory campaign, 64 new simulations and approximately 1.404 × 10ˆ9 additional encounters were generated under prespecified criteria. Directional alignment with the mean-field latent tangent was supported at all four tested population sizes, with median cosines between 0.984

Table 1
Prospective confirmatory evaluation of the reduced finite-population flow.

Population N	Flow cosine (simultaneous interval)	Clock mismatch (simultaneous interval)	Median transverse fraction	Primary interpretation
25,000	0.9863 [0.9755, 0.9929]	0.0495 [0.0333, 0.0775]	0.1649	Direction supported; clock agreement supported
75,000	0.9841 [0.9767, 0.9891]	0.0663 [0.0413, 0.0841]	0.1779	Direction supported; clock agreement supported
250,000	0.9894 [0.9832, 0.9940]	0.0620 [0.0358, 0.0753]	0.1454	Direction supported; clock agreement supported
1,000,000	0.9937 [0.9869, 0.9976]	0.0608 [0.0229, 0.0766]	0.1125	Direction supported; clock agreement supported

Note. Flow direction was supported when the simultaneous interval remained above 0.95. Clock agreement was supported when the simultaneous interval for |alpha_N - dphi_N/dtau| remained below 0.10. Strict monotonic decline of the transverse fraction across adjacent population sizes was not established; all three Bonferroni-adjusted pairwise contrasts were indeterminate.

and 0.994 and simultaneous intervals above 0.95. The independently estimated tangent speed and distributional phase speed also agreed within the prespecified tolerance of 0.10. These results, summarized in Figs. 1 and 2 and Table 1, provide the strongest evidence for the representation $\dot{z_N} \approx v_N G_\infty(z_N) + r_N^\perp$: finite populations predominantly move in the same reduced direction as mean-field but traverse that flow more slowly.

The stronger composite hypothesis was not fully confirmed because strict pairwise monotonic decline of the transverse component with N remained statistically indeterminate (Fig. 3). This distinction is important. The appropriate conclusion is neither that the entire framework was confirmed nor that it was rejected. The common-flow direction and effective-clock agreement received prospective support; the precise asymptotic scaling of the transverse correction remains unresolved. Maintaining that separation avoids post hoc relaxation of the confirmatory criterion while preserving the results that genuinely survived prospective testing. The increase from N = 25,000 to N = 75,000 may reflect seed-to-seed variability with only 16 independent seeds per group, as well as sensitivity of the phase matching, smoothed derivatives and transverse-to-total velocity ratio. An intermediate-scale interaction between demographic fluctuations and clipping is another possible explanation, but the present comparisons do not distinguish these mechanisms. The non-monotonicity therefore establishes neither a crossover nor a sampling artifact; resolving its origin requires targeted replication or boundary-treatment sensitivity tests, separate from the frozen confirmatory analysis.

Taken together, the findings refine rather than contradict the first article. The original cooperation values remain genuine properties of the simulation protocol, but terms such as "critical region", "collapse", and "equilibrium" require a more careful dynamical interpretation under the audited Fermi rules. The first paper discovered the macroscopic regularity; the present study identifies the evolutionary clock, the distributional state, the finite-size slowdown, and the reduced flow that generate it. More broadly, the results illustrate a general modeling caution: when a parameter changes the number of microscopic events per macroscopic simulation step, apparent parameter effects may partly reflect unequal dynamical exposure rather than a change in the underlying microscopic mechanism.

6. Conclusion

This study provides the mathematical counterpart to the interaction-density phenomenon reported by De los Reyes-Marín et al. [1]. In the audited sequential Fermi process, interaction density does not alter the one-encounter transition rule; it maps nominal generations into the natural evolutionary clock $\tau = 2gI/N$. For positive mutation, the stationary finite-population distribution is therefore independent of I, and the cooperation deterioration observed at large interaction density is best understood as a finite-horizon relaxation effect rather than an I-induced stationary bifurcation.

The microscopic process also admits an exact payoff reduction through $a = \tanh[\beta(T-S)/2]$, and the finite-population drift shows that mean cooperation does not close the dynamics. In the mutation-free mean-field limit, the explicit solution $C(\tau) = 1/(a\tau) - 1/(e^{a\tau} - 1)$ explains the plateau-and-decline structure analytically. With positive mutation, the state is genuinely distributional: finite populations traverse the mean-field dynamics more slowly because higher-order features of the strategy distribution develop differently, producing a hierarchy from variance to skewness and the fourth cumulant.

A reduced distributional description captures the dominant geometry of this process. Finite populations evolve mainly along the same latent mean-field flow at an effective speed v_N, with a smaller transverse correction. Prospective simulations with frozen criteria supported both common-flow direction and agreement between independent effective-speed estimates, while strict monotonic decay of the transverse component remained unresolved. The resulting framework is therefore $I \to \tau \to f(p, \tau) \to z(\tau) \to C(\tau)$, with finite-size dynamics approximated by $\dot{z_N} \approx v_N G_\infty(z_N) + r_N^\perp$. Within the audited model, this structure explains why interaction density appeared to dominate cooperation in the original simulations and clarifies the distinction between structural evolutionary effects and changes in evolutionary exposure.

7. Limitations

The conclusions are specific to the audited sequential Fermi implementation. The published narrative and revised executable model are not fully aligned with respect to mutation timing, so the exact clock, drift identities, and mean-field equations derived here should not be transferred automatically to the alternative verbal mutation rule. Likewise, the stationary-distribution result requires positive mutation and finite selection intensity; the mutation-free process has qualitatively different absorbing behavior.

The interaction-clock identity is exact only because I changes the number of applications of an otherwise unchanged encounter kernel. If interaction density altered partner selection, accumulated payoffs, simultaneous updating, network topology, information, or congestion at the microscopic level, it could become a genuinely structural parameter rather than a clock. The mean-field equation is also a deterministic large-population reference rather than a complete theorem of finite-population convergence: no quantitative propagation-of-chaos bound uniform in time is established here.

The explicit closed-form solution applies only to the mutation-free mean-field equation with uniform initialization. Quantitative finite-size analyses were concentrated on the baseline Prisoner's Dilemma, $\mu = 0.02$, $\beta = 1$, and initially cooperative populations, so the magnitude of slowdown and the reduced-flow geometry should not be treated as universal across mutation strengths, initial conditions, games, or update rules. The moment closures and POD model are likewise approximations: the failure of low-order closures shows that a few conventional moments are insufficient, while four POD modes describe the dominant trajectory but do not prove the existence of a universal four-dimensional invariant manifold.

The phase-based clock is an estimator obtained through Hellinger matching, monotonicity enforcement, histogram discretization, and smoothing; it is not an independently observed physical time. Its credibility is strengthened by agreement with tangent-projection speeds, but methodological dependence remains. The confirmatory protocol was prospectively frozen and SHA-256 sealed rather than externally preregistered. The confirmatory campaign used four population sizes and 16 seeds per size: this was sufficient for the prespecified directional and clock tests, but not for an asymptotic law. The composite hypothesis was not fully confirmed because strict pairwise monotonic decline of the transverse component remained indeterminate. The terminal-bin aggregation also limits the resolution of clipping-induced boundary mass; the reduced representation does not provide a separate dynamical coordinate for each Dirac atom.

Finally, encounter counts are not independent statistical observations; simulation seeds are the relevant independent units. The model also excludes networks, persistent partners, memory, reputation, institutions, and dynamic payoffs. These restrictions are deliberate because they isolate the clock mechanism, but they limit external generalization. Future work should derive formal convergence rates, test alternative mutation and selection regimes, and introduce structured interaction mechanisms to determine which parts of the clock–flow decomposition persist beyond the present model.

CRediT authorship contribution statement

Óscar De los Reyes-Marín: Writing – original draft, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Iria Paz Gil:** Writing – review & editing, Methodology, Data curation. **Jose Torres-Pruñonosa:** Visualization, Validation, Supervision. **Raúl Gómez-Martínez:** Visualization, Validation, Supervision.

Declaration of competing interest

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

Appendix. Distributional modes and exploratory diagnostics

The following panels display the original saved POD basis and exploratory diagnostic data; the reduced model and confirmatory criteria are unchanged.

At the normalized squared mean square-root distribution, increasing mode 1 lowers mean cooperation and increases spread; mode 2 primarily broadens the distribution with a smaller mean shift. Modes 3 and 4 have opposing effects on the high-cooperation terminal bin and additional asymmetric interior structure. These interpretations are local to the reference distribution and basis orientation. The modes are not pure mean, variance, skewness or boundary coordinates.

Data availability

All data, results, and software required to reproduce the study have been provided as Supplementary Material accompanying this submission.

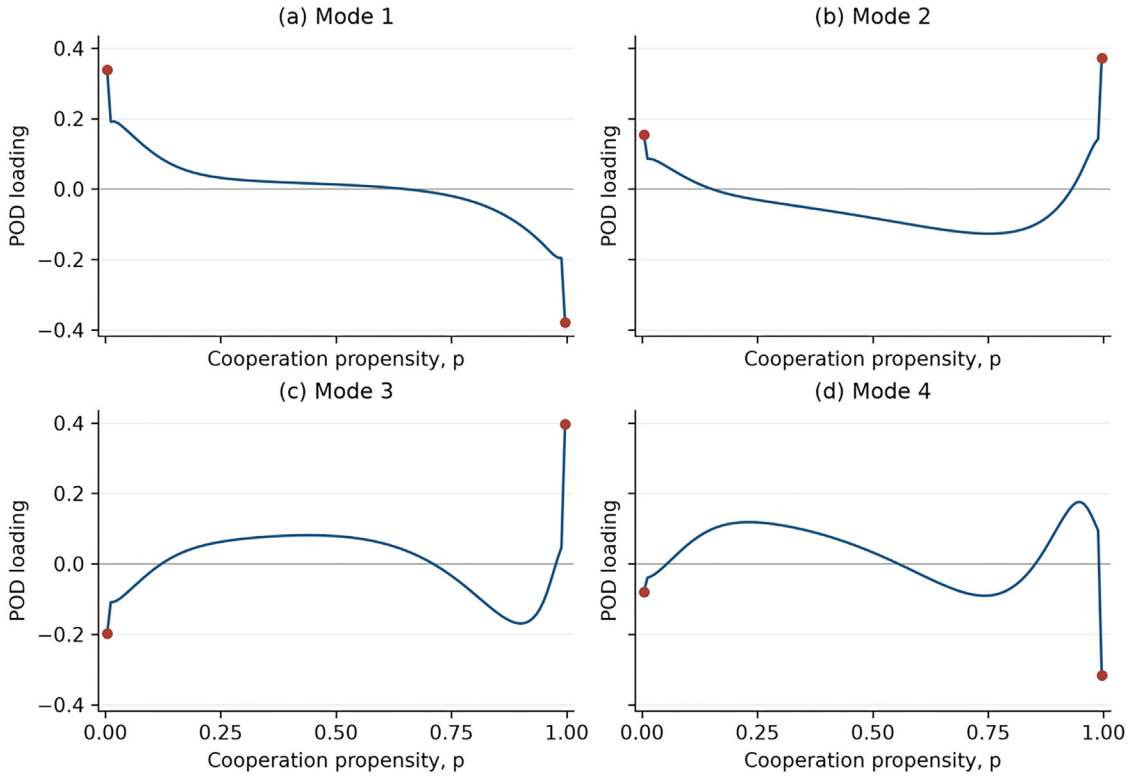


Fig. A.1. Spatial shapes of the four retained POD modes in square-root probability coordinates. The basis is evaluated on the original 128 equal-width bins, with the stored sign convention. Red markers identify the terminal bins containing the boundary atoms and nearby interior mass. The modes individually explain 61.26%, 24.89%, 9.22% and 3.24% of trajectory variation (98.62% cumulatively).

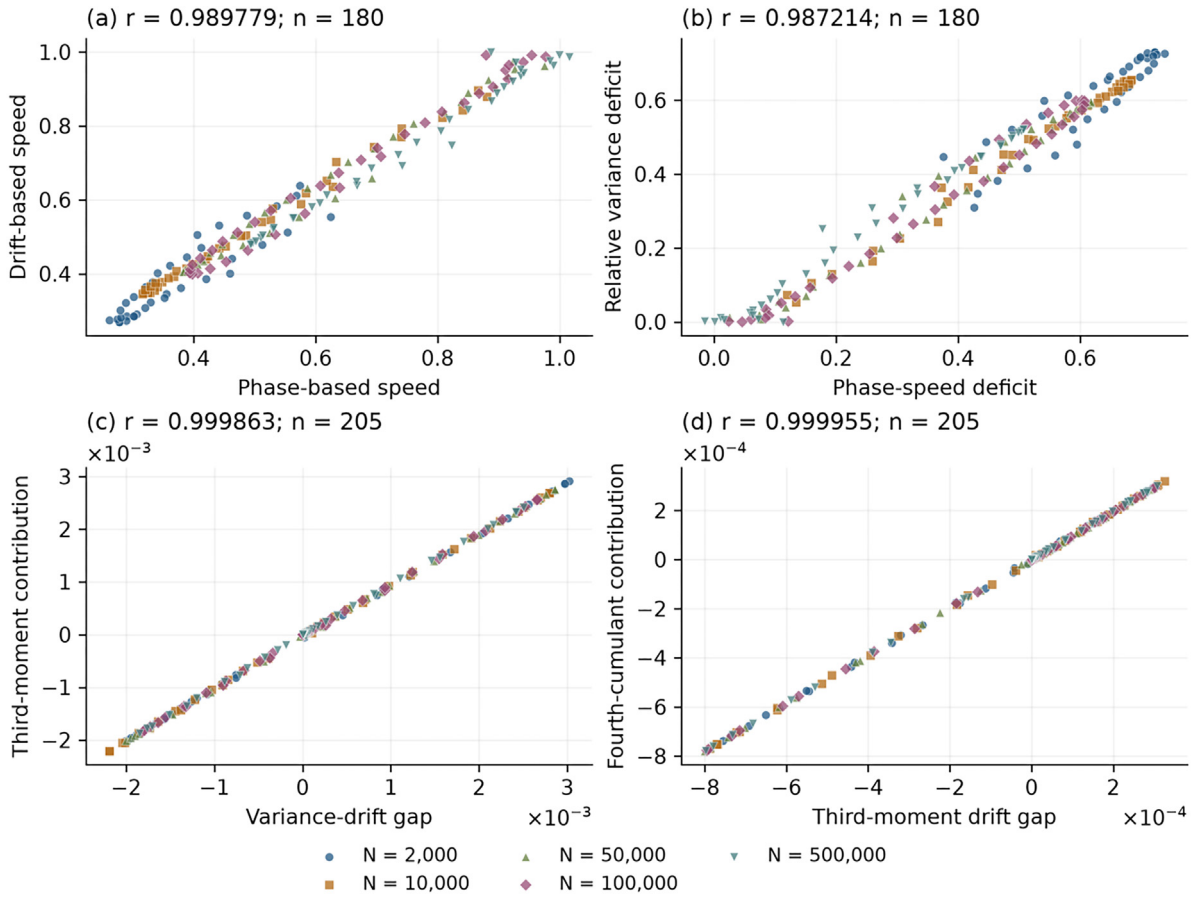


Fig. A.2. Exploratory diagnostics reconstructed from the saved analysis data. (a) Effective speed from the mean-cooperation phase estimator versus the exact-drift ratio. (b) Phase-speed deficit versus the relative variance deficit at matched cooperation. (c) Variance-drift gap versus its third-central-moment contribution. (d) Third-moment drift gap versus its fourth-cumulant contribution. Panels (a–b) use $50 \leq \tau \leq 120$ (180 population-size/time summaries); panels (c–d) use $40 \leq \tau \leq 120$ (205 summaries). Colors and symbols distinguish the five population sizes. Reported r values are descriptive Pearson correlations across these summaries; temporal observations are dependent and do not constitute additional independent seeds. Panels (c–d) compare a total with a component of that total and therefore document the decomposition rather than an independent causal test.

References

[1] De los Reyes-Marín Ó, Paz Gil I, Torres-Prunonosa J, Gómez-Martínez R. Interaction density drives cooperation in large evolutionary populations: Evidence from large-scale agent-based simulations. Chaos Solitons Fractals 2026;212:119178. http://dx.doi.org/10.1016/j.chaos.2026.119178.

[2] Maynard Smith J. Evolution and the theory of games. Cambridge University Press; 1982, http://dx.doi.org/10.1017/CBO9780511806292.

[3] Hofbauer J, Sigmund K. Evolutionary games and population dynamics. Cambridge University Press; 1998, http://dx.doi.org/10.1017/CBO9781139173179.

[4] Nowak MA. Evolutionary dynamics: exploring the equations of life. Harvard University Press; 2006.

[5] Axelrod R. The evolution of cooperation. Basic Books; 1984.

[6] Hamilton WD. The genetical evolution of social behaviour. I. J Theoret Biol 1964;7(1):1–16. http://dx.doi.org/10.1016/0022-5193(64)90038-4.

[7] Nowak MA, May RM. Evolutionary games and spatial chaos. Nature 1992;359(6398):826–9. http://dx.doi.org/10.1038/359826a0.

[8] Nowak MA, Sigmund K. Evolution of indirect reciprocity by image scoring. Nature 1998;393(6685):573–7. http://dx.doi.org/10.1038/31225.

[9] Perc M, Gómez-Gardeñes J, Szolnoki A, Floría LM, Moreno Y. Evolutionary dynamics of group interactions on structured populations: A review. J R Soc Interface 2013;10(80):20120997. http://dx.doi.org/10.1098/rsif.2012.0997.

[10] Traulsen A, Nowak MA, Pacheco JM. Stochastic dynamics of invasion and fixation. Phys Rev E 2006;74(1):011909. http://dx.doi.org/10.1103/PhysRevE.74.011909.

[11] Ruijgrok M, Ruijgrok TW. An effective replicator equation for games with a continuous strategy set. Dyn Games Appl 2015;5(2):157–79. http://dx.doi.org/10.1007/s13235-014-0118-y.

[12] Karev GP. How to explore replicator equations?. 2008, http://dx.doi.org/10.48550/arXiv.0812.4295, arXiv.

[13] Vardanyan E, Saakian DB. The analytical dynamics of the finite population evolution games. Phys A 2020;553:124233. http://dx.doi.org/10.1016/j.physa.2020.124233.

[14] Cheng K, Zimmermann R. Adaptive data-driven probabilistic reduced-order models for parameterized dynamical systems. SIAM J Sci Comput 2025;47(2):C426–50. http://dx.doi.org/10.1137/23M1602280.

[15] Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. Rev Modern Phys 2009;81(2):591–646. http://dx.doi.org/10.1103/RevModPhys.81.591.
[16] Epstein JM. Generative social science: studies in agent-based computational modeling. Princeton University Press; 2006.
[17] Perc M. The social physics collective. Sci Rep 2019;9:16549. http://dx.doi.org/10.1038/s41598-019-53300-4.
[18] Schmidt M, Lipson H. Distilling free-form natural laws from experimental data. Science 2009;324(5923):81–5. http://dx.doi.org/10.1126/science.1165893.
[19] Price GR. Selection and covariance. Nature 1970;227:520–1. http://dx.doi.org/10.1038/227520a0.
[20] Frank SA. The price equation, Fisher's fundamental theorem, kin selection, and causal analysis. Evolution 1997;51(6):1712–29, https://stevefrank.org/reprints-pdf/97Evol-Causal.pdf.