From Bellman to Boltzmann: Towards a Kinetic Theory of Controlled Emergence and Meta-Cognition

Jia Liu[1], **Alvaro Velasquez**[2], **Susmit Jha**[3], **Nathaniel D. Bastian**[4]

[1]ECE Department, The Ohio State University
[2]CS Department, University of Colorado
[3]Defense Advanced Research Projects Agency
[4]Pathfinder Applied Research, LLC

Abstract

We propose a research vision for a kinetic theory of collective intelligence in large-scale multi-agent learning systems. Contemporary multi-agent reinforcement learning (MARL) provides powerful tools for optimizing individual agents, while mean-field game theory characterizes population-level decision-making in the large population limit. Yet, these frameworks offer limited mathematical understanding of how higher-order collective structures (e.g., coordination, specialization, adaptive communication, and meta-cognition) emerge from microscopic interactions among heterogeneous learning agents. Inspired by statistical mechanics, we argue that a mesoscopic kinetic description can provide a missing bridge between individual learning dynamics and macroscopic collective behavior. Building on existing connections between Boltzmann-type equations and mean-field games, we envision Bellman–Boltzmann dynamics in which Bellman decision making is coupled with Boltzmann-type interaction dynamics over heterogeneous agent states, including policies, beliefs, memory, and other cognitive variables. This perspective raises fundamental questions about the well-posedness and mean-field limits of such systems, entropy and Lyapunov principles for learning dynamics, phase transitions in collective behavior, and the emergence and control of collective meta-cognition. In this paper, we outline an ambitious mathematical research program spanning microscopic agent models, kinetic equations, macroscopic limits, and controllable emergence and meta-cognition, together with algorithmic and experimental directions for large-scale MARL and multi-agent AI. Our central thesis is that collective intelligence should be studied not only as an optimization problem, but as an emergent dynamical phenomenon governed by principles analogous to those of statistical mechanics.

1 Introduction

Background and Motivation: Multi-agent systems are entering a regime in which intelligence is increasingly distributed across large populations of interacting agents. Advances in multi-agent reinforcement learning (MARL), foundation models, and autonomous systems have enabled agents to coordinate, communicate, specialize, and solve tasks that are difficult for an individual agent. Examples range from robot swarms and autonomous vehicles to large-scale language-model-based agent societies and distributed AI systems. As the number, heterogeneity, and autonomy of agents increase, however, a fundamental challenge is to understand how collective intelligence and cognition emerge from the interactions among individual learning agents.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

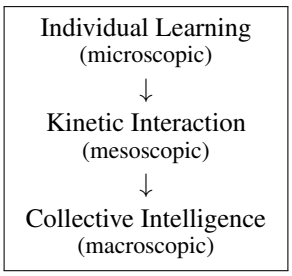
Existing approaches largely address this question from an optimization perspective. MARL seeks policies that maximize individual or collective rewards, while mean-field games (MFGs) and mean-field control replace explicit interactions among a large number of agents by their population distribution. These frameworks provide powerful tools for analyzing small and large populations, but they generally do not provide a kinetic description of *how* information, strategies, beliefs, and capabilities propagate through the growth of a population. In particular, the transition from microscopic learning and decision-making to mesoscopic interaction dynamics and macroscopic collective organization remains poorly understood. This missing mesoscopic layer becomes important when agents are heterogeneous, adaptive, and capable of changing not only their actions but also their policies, beliefs, memory, and interaction strategies.

Statistical mechanics provides a natural conceptual blueprint for addressing this gap. Kinetic theory bridges microscopic particle dynamics and macroscopic physical laws by describing the evolution of a population distribution under transport and interaction processes. The Boltzmann equation is particularly powerful because it does not require tracking every microscopic particle while retaining information about pairwise interactions. This perspective suggests a fundamental question for artificial intelligence:

Question: Can a kinetic theory play an analogous role for populations of learning agents?

Our Vision: In this paper, our vision is to develop a *kinetic theory of heterogeneous learning agents and collective intelligence through the lens of a Bellman-Boltzmann system*. We seek to extend the state, interaction, and objective structures of conventional MARL systems and knowledge-evolution models to general learning agents whose internal cognitive states evolve through interaction. At the microscopic level, we envision an agent's *"cognitive state"* that captures its environment state, policy, belief or world model, memory or meta-state, and heterogeneous capabilities. Individual decision-making is governed by a Bellman prin-

ciple, while the population distribution evolves through a Boltzmann-type kinetic equation, which describes interactions such as communication, imitation, knowledge transfer, cooperation, competition, and specialization among the agents. The essential conceptual transition is therefore



We envision three particularly important features of the proposed Bellman-Boltzmann framework:

- *1) Heterogeneity Becomes a First-Class Object of the Kinetic Theory:* Rather than representing agents by a single knowledge variable, the kinetic state encodes heterogeneous policies, capabilities, beliefs, memories, and learning mechanisms. This makes it possible to study whether and how fast useful population-level structures (e.g., specialization, leadership, role differentiation, and division of cognitive labor) can emerge spontaneously from local interactions.
- *2) Emergence Becomes a Mathematical Evolution Rather Than Merely an Empirical Observation:* We envision defining macroscopic order parameters for properties such as coordination, specialization, collective competence, or communication efficiency. Changes in these order parameters could then be analyzed through stability, bifurcation, and phase-transition theory. The fundamental question becomes not simply whether agents learn a good policy, but *under what conditions a population undergoes a transition from disorganized individual behavior to organized collective intelligence*.
- *3) Collective Meta-Cognition as a Self-Referential Kinetic Phenomenon:* Let $C_t := \mathcal{C}(f_t)$ represent a macroscopic measure of the distribution f_t of the population's uncertainty, competence, coordination, or learning state. If this variable feeds back into the Boltzmann interaction dynamics, then the population can potentially monitor and regulate aspects of its own learning process. This creates the feedback loop:

  $$f_t \to C_t \to \text{Interactions conditioned on } C_t \to f_t.$$

  This enables the investigation on *meta-cognition* in a *dynamical* fashion: the population develops information about its own collective cognitive state and uses that information to modify its subsequent behavior. This provides a potential mathematical mechanism and theoretical foundation for studying meta-cognition in higher-order collective intelligence without being limited to assumptions in which every individual agent contains a global and static meta-cognitive module.

Technical Challenges: Realizing our proposed vision is highly non-trivial and requires overcoming *four* substantial mathematical and computational challenges. The first challenge is the construction of an appropriate microscopic model that simultaneously retains the Markov decision-process structure of reinforcement learning and admits a tractable kinetic limit. The second challenge stems from the derivation and analysis of kinetic equations when interactions modify not only physical states but also policies, beliefs, memories, and other adaptive variables. The third challenge is to establish rigorous connections between finite-agent MARL systems and their kinetic limits, including propagation-of-chaos, stability, and approximation results. Last and perhaps most importantly, the fourth challenge lies in identifying entropy or Lyapunov principles that characterize the direction of collective learning and to determine whether analogues of the classical H-theorem can be established. Collectively, we need to understand when collective order emerges, how it can be detected mathematically, and whether it can be deliberately induced or suppressed.

We envision addressing these challenges through a *hierarchy* of models, beginning with analytically tractable microscopic learning systems and progressively introducing heterogeneous capabilities, adaptive interactions, and richer cognitive states. This program will combine kinetic PDE theory, stochastic control, mean-field games, optimal transport, dynamical-systems theory, and modern learning theory. Rather than attempting to model the full complexity of contemporary AI from the outset, the goal is to identify minimal microscopic models that preserve the essential mechanisms of learning and interaction, thereby exposing general principles of collective intelligence.

Research Impacts: If successful, this research could establish a new mathematical connection between statistical mechanics, mean-field games, control, and artificial intelligence. Scientifically, our proposed research will provide a theoretical framework for determining whether concepts such as entropy production, phase transitions, symmetry breaking, and free-energy minimization have meaningful counterparts in adaptive multi-agent learning populations. It could also lead to rigorous criteria for the emergence of coordination, specialization, and collective meta-cognition.

The technological implications are equally broad. A kinetic perspective could lead to new algorithms that optimize not only individual agent policies but also the *population-level dynamics of learning*. Such methods could potentially improve the scalability and robustness of robot swarms, autonomous transportation systems, distributed edge intelligence, multi-agent scientific discovery, and societies of language-model-based agents. More fundamentally, the proposed framework seeks to transform emergence from a phenomenon that is primarily observed after the fact into a property that can be *predicted, designed, and controlled*.

The remainder of this paper is organized as follows: Section 2 surveys the most related research areas, putting our work in comparison perspectives. Section 3 reviews and draws the relevant lessons from statistical mechanics and kinetic theory. Section 4 introduces the proposed Bellman-Boltzmann framework for heterogeneous learning in multi-agent systems. Section 5 investigates controlled emergence, phase transitions, and collective meta-cognition. Section 6

outlines the principal mathematical challenges and a research program toward rigorous mathematical foundations. Finally, Section 7 discusses algorithmic realization, experimental validation, applications, and the broader vision of a science of collective intelligence.

2 Related Work

To our knowledge, the basic mathematical idea of coupling a Boltzmann-type population equation with a Bellman-type optimal-control equation for multi-agent systems historically traces its roots to the model of Lucas and Moll (Lucas and Moll 2014) for knowledge-driven economic growth, where individuals acquire knowledge through interactions with more knowledgeable individuals and choose how to allocate time between learning and production. Subsequently, Burger, Lorz, and Wolfram formulated and mathematically analyzed this model as a *Boltzmann mean-field game* (BMFG), which couples a Boltzmann-type equation describing the evolution of the population density with respect to knowledge level to a Hamilton–Jacobi–Bellman (HJB) equation describing the optimal learning strategy. They established the local-in-time existence and uniqueness for the coupled system and analyzed balanced-growth-path solutions (Burger, Lorz, and Wolfram 2016). Burger, Lorz, and Wolfram further studied balanced-growth-path solutions of this BMFG model, showing that suitable initial distributions can lead to exponential growth of aggregate production and investigating extensions in which knowledge also evolves through stochastic dynamics (Burger, Lorz, and Wolfram 2017). More recently, Burger, Kanzler, and Wolfram generalized the BMFG framework to incorporate limits to learning and more general utility functions. Their formulation explicitly couples a Boltzmann equation for the distribution of knowledge with a Bellman equation determining the optimal fraction of time devoted to learning, resulting in a nonlinear forward–backward PDE system (Burger, Kanzler, and Wolfram 2024).

The aforementioned works establish an important precedent: *Boltzmann-type population dynamics and Bellman optimal decision-making can indeed be coupled into a mathematically analyzable mean-field model*. They also demonstrate the value of kinetic theory for populations in which interactions fundamentally change the distribution of agent states. More broadly, kinetic equations have been used to model large interacting populations in economics and social systems, including wealth distributions, opinion formation, ranking, and other collective phenomena; kinetic-game formulations have also been developed for domains such as insurance markets (Brinkman and Ringhofer 2017). Nevertheless, these existing models leave *a substantial conceptual gap from the perspective of modern AI.* First, their microscopic state is typically a *low-dimensional physical, economic, or knowledge variable*, rather than the rich internal state of a learning agent. In the knowledge-growth BMFG (Burger, Lorz, and Wolfram 2016; Burger, Kanzler, and Wolfram 2024), for example, agents are characterized primarily by their knowledge level, and interactions modify that scalar state. The Bellman component determines how much time an agent optimally allocates to learning, rather than how a general reinforcement-learning agent learns a policy, maintains beliefs, or changes its decision-making mechanism.

Second, existing BMFGs primarily study population evolution toward economically meaningful outcomes, such as balanced growth, *rather than the emergence* of qualitatively new collective cognitive structures. Their interaction operators describe knowledge exchange, but do not explicitly model the emergence of *coordination, specialization, communication protocols, collective memory, or higher-order reasoning.* Third, although the underlying Boltzmann framework naturally suggests questions about entropy, equilibrium, and phase transitions, existing BMFG analyses have *not* developed a general theory in which such phenomena correspond to the emergence of *collective intelligence*. Finally, these models do not provide a mechanism through which a population can form a representation of its *own collective cognitive state* and use that representation to regulate its future learning dynamics.

In contrast, the key difference of our proposed research from existing BMFG models is consequently *not the mere presence of a Bellman equation and a Boltzmann equation*. Rather, it is the proposed use of kinetic theory to study the dynamics of **adaptive cognitive states** and the emergence of qualitatively new collective structures.

We note that there is also a broader distinction between our proposed vision and recent learning-based MFG research. For example, learning mean-field games have combined Q-learning, Boltzmann policies, and mean-field equilibrium computation for large-population stochastic games (Guo et al. 2019). Here, however, "Boltzmann" refers to a Boltzmann policy or policy distribution rather than a Boltzmann kinetic equation. The objective in (Guo et al. 2019) remains learning an equilibrium policy rather than modeling the kinetic evolution of heterogeneous cognitive states. Similarly, recent mathematical work on kinetic-type MFGs studies forward-backward systems associated with controlled kinetic dynamics, including agents with spatial and velocity variables (Ambrose, Griffin-Pickering, and Mészáros 2024). These developments provide important mathematical foundations, but they do not address the emergence of higher-order collective cognitive behavior in learning populations.

3 Lessons from Statistical Mechanics

Statistical mechanics provides a useful conceptual blueprint for understanding collective behavior in large populations. Its central challenge is closely related to that of large-scale multi-agent learning: how can macroscopic behavior emerge from interactions among a vast number of heterogeneous microscopic entities? Kinetic theory addresses this question by introducing an intermediate, mesoscopic description between microscopic dynamics and macroscopic laws. Rather than tracking every particle individually, it describes the evolution of a population distribution while retaining the essential effects of microscopic interactions. The Boltzmann equation is the canonical example of this approach (Boltzmann 1872; Saint-Raymond 2009). This suggests the following hierarchy we envision for learning populations:

1) Individual Learning (Microscopic) → 2) Kinetic Interaction (Mesoscopic) → 3) Collective Intelligence (Macroscopic). Specifically, the microscopic level describes individual decision-making and learning; the kinetic level describes how interactions reshape the population distribution; and the macroscopic level characterizes emergent collective properties. In what follows, we highlight *three lessons* from statistical mechanics that motivate this perspective:

Lesson 1) Population Distributions Preserve Interaction Structure. For a many-particle system, the microscopic state $Z_N = (z_1, \ldots, z_N)$ becomes intractable as N grows. Kinetic theory instead introduces a distribution $f(t, z)$ describing the population density over the state space. For a dilute gas, the Boltzmann equation takes the form:

$$\partial_t f + v \cdot \nabla_x f = Q(f, f),$$

where the collision operator Q captures the effect of *pairwise* interactions. The important lesson for learning systems is that a population-level equation can retain information about microscopic interactions without explicitly tracking every agent. In a learning population, interactions may correspond to *communication, imitation, knowledge transfer, cooperation, competition, or specialization.* This motivates a kinetic learning equation of the generic form:

$$\partial_t f + \mathcal{L}_{\mathrm{MDP}} f = Q(f, f),$$

where operator $\mathcal{L}_{\mathrm{MDP}}$ captures individual interaction with the environment and Q captures interactions among agents (both to be defined later in Section 4).

Lesson 2) Macroscopic Behavior Emerges from Kinetic Dynamics. A second lesson is that the kinetic distribution is itself an intermediate description. In physical systems, macroscopic quantities such as density, momentum, and energy are moments of f, and appropriate scaling limits can lead from the Boltzmann equation to macroscopic fluid equations (Saint-Raymond 2009). This suggests seeking analogous macroscopic observables for learning populations. Toward this end, let $M(f)$ denote an order parameter measuring a population-level property such as *coordination, specialization, collective competence, or communication efficiency.* The emergence of collective intelligence can then be formulated as a transition in the behavior of $M(f_t)$, rather than solely as the optimization of individual-agent policies. Importantly, such collective structures may not be explicitly represented by any individual agent and may only become visible at the population level.

Lesson 3) Entropy Provides an Organizing Principle. The Boltzmann equation offers a profound example of how a population-level functional can characterize collective evolution. The Boltzmann H functional is defined as:

$$H[f] = \int f \log f,$$

and the celebrated H-theorem establishes, under the classical assumptions, that:

$$\frac{d}{dt} H[f] \leq 0.$$

Thus, microscopic interactions induce a **macroscopic directionality in time** and provide a Lyapunov-like principle for the approach toward equilibrium (Boltzmann 1872).

For multi-agent learning populations, it is unlikely that physical entropy itself will provide the appropriate quantity: learning agents are open, adaptive systems driven by rewards, information acquisition, and environmental interactions. Nevertheless, the H-theorem suggests searching for "generalized entropy," "free-energy," or Lyapunov functionals that characterize the evolution of learning populations. Establishing whether such principles exist, and what they imply about convergence, stability, and emergence, is a central open question of our research program.

Collectively, the aforementioned three lessons motivate our statistical-mechanical view of collective intelligence. We note that the *kinetic level* is particularly important because it provides a **bridge** between individual learning and population-level emergence: a purely microscopic description becomes intractable at scale, while a purely macroscopic description may discard the interaction mechanisms responsible for emergence and meta-cognition. The next section builds on this perspective by coupling Bellman optimality at the agent level with Boltzmann-type population dynamics, leading to the proposed Bellman–Boltzmann framework for heterogeneous learning agents.

4 The Bellman-Boltzmann System for Multi-Agent Intelligence

We now develop the central mathematical object of our vision: a *Bellman-Boltzmann* system for a large population of interacting learning agents. The key idea is to combine *two complementary descriptions* of intelligence: At the *microscopic level,* each agent is a reinforcement learner that makes sequential decisions according to Bellman optimality. At the *mesoscopic level,* agents interact with one another, causing their internal states to evolve according to Boltzmann-type population dynamics. The resulting system provides a bridge between individual decision making and collective behavior.

Toward this end, consider a population of N agents interacting with an environment in a *reinforcement learning* fashion, which is modeled as a Markov decision process (MDP). For a conventional agent, the state and action are denoted by (s, a), and its policy $\pi(a|s)$ determines the probability of taking action a in state s. However, we argue that, for a kinetic description, (s, a) alone is *insufficient* to characterize the state of a learning agent: two agents occupying the same environmental state may have very different policies, beliefs, memories, or learning capabilities. Consequently, they respond very differently to the same observation. We therefore introduce the following augmented adaptive cognitive state at each agent (omitting the agent index i for brevity):

$$z = (s, \pi, b, m, \theta),$$

where s denotes the environmental state, π denotes the agent's policy, b represents its belief or internal representation of the environment, m denotes its memory or internal learning state, and θ is a parameter that captures the persistent heterogeneity across agents. Then, the microscopic

population state at time t can be written as:

$$Z_N(t) = \{z_1(t), \ldots, z_N(t)\},$$

while the corresponding kinetic description is a population distribution $f_t(z)$ over this extended state space.

The Microsopic Level: The individual learning dynamics at the microsopic level are governed by the *Bellman optimality* principle similar to conventional MARL systems: Given a population distribution f, let $V_f(z)$ denote the value of an agent in state z. Then, its optimal value satisfies

$$V_f(z) = \mathcal{T}_f V_f(z),$$

where $\mathcal{T}_f$ is the Bellman operator induced by the environment and the population state. In the standard MDP formulation, this operator takes the form:

$$V(s) = \max_{a \in \mathcal{A}} \left\{ r(s, a; f) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a,f)}[V(s')] \right\}.$$

The dependence on f emphasizes that the reward, transition law, or effective decision problem of an agent may depend on the behavior of the population. Thus, Bellman optimality provides the microscopic decision-making principle, while the population distribution f determines the strategic environment in which that decision is made.

The Mesoscopic Level: The second component of the Bellman-Boltzmann system is the kinetic evolution of the population at the mesoscopic level. In analogy with the Boltzmann equation, we have

$$\partial_t f + \mathcal{L}_{\mathrm{MDP}} f = Q(f, f),$$

where $\mathcal{L}_{\mathrm{MDP}}$ represents the evolution of individual agents through their interaction with the environment and $Q(f, f)$ represents interactions among agents. The operator Q is the key mechanism through which collective learning occurs. Depending on the application, an interaction may represent communication, imitation, knowledge transfer, cooperation, competition, or specialization. In contrast to physical collisions, these interactions can change the internal learning state of an agent: for example, an agent may update its policy after observing another agent, revise its belief after receiving information, or alter its memory through shared experience.

The combination of the Bellman and kinetic descriptions gives the generic Bellman-Boltzmann system:

$$\boxed{\begin{aligned} V_f &= \mathcal{T}_f V_f, \\ \partial_t f + \mathcal{L}_{\mathrm{MDP}} f &= Q(f, f). \end{aligned}}$$

The two equations are coupled in both directions. The Bellman equation depends on the population distribution through $\mathcal{T}_f$, because an agent's optimal decision depends on the behavior of other agents. Conversely, the kinetic evolution depends on the decisions generated by the Bellman equation through $\mathcal{L}_{\mathrm{MDP}}$ and potentially through the interaction operator Q. Consequently, the system is inherently forward–backward: individual optimality shapes population evolution, while population evolution continuously changes the decision problem faced by each agent.

This formulation also provides a natural way to incorporate heterogeneity. The parameter θ can represent differences in capabilities, learning rates, preferences, computational resources, or other persistent agent characteristics. The population distribution then takes the form:

$$f_t = f_t(s, \pi, b, m, \theta),$$

allowing the theory to describe not only the evolution of the average agent but also the evolution of the entire distribution of heterogeneous learning capabilities and cognitive states. In particular, interactions can reshape the distribution itself, potentially producing concentration, dispersion, clustering, or specialization in different regions of the state space.

The Bellman-Boltzmann system therefore differs fundamentally from simply applying a mean-field approximation to MARL. A conventional mean-field description replaces the detailed interaction among agents by a population statistic, whereas the kinetic formulation retains an explicit description of how pairwise interactions transform the distribution of agent states. This additional level of resolution is essential for studying emergence: macroscopic structures such as coordination or specialization may depend not only on the population average but on the detailed interaction mechanisms that produce changes in the population distribution.

We note that the proposed Bellman-Boltzmann system should be viewed as a general *meta-framework* rather than a single fixed model. Different choices of agent state, interaction kernel, Bellman operator, and scaling assumptions produce different Bellman-Boltzmann systems. The fundamental research problem is to identify classes of such systems that are both sufficiently expressive to capture meaningful multi-agent learning phenomena and sufficiently structured to admit rigorous analysis. In the remainder of the paper, we focus on the collective phenomena that such a framework may make accessible, particularly the controlled emergence and organized population-level meta-coginition.

5 Controlled Emergence and Meta-Cognition

The central promise of the Bellman–Boltzmann framework is not merely to describe a large population of learning agents, but to provide a mathematical language for understanding how *collective intelligence emerges*. In conventional MARL, emergence is often evaluated after training by examining the behavior of the resulting agents. A kinetic perspective offers a different viewpoint: emergence can be studied as a dynamical transition in the population distribution. This raises two questions: *when does organized collective behavior emerge, and can this emergence be deliberately induced or controlled?* We envision addressing these questions by connecting kinetic population dynamics with macroscopic order parameters and feedback mechanisms.

Emergence: Population-Level Phase Transitions

Let $f_t(z)$ denote the population distribution over the extended agent state introduced in Section 4. Although the individual dynamics may be high-dimensional and heterogeneous, collective behavior can often be characterized through a lower-dimensional functional $M(f_t)$, which we

refer to as a *macroscopic intelligence order parameter*. Depending on the application, $M(f)$ may measure coordination, specialization, collective competence, communication efficiency, diversity, or any population-level property.

Our key idea is that qualitative changes in collective behavior can then be formulated as *transitions in the order parameter.* Let λ denote a population-level control parameter, such as interaction strength, information exchange rate, communication bandwidth, exploration level, or reward coupling. The kinetic dynamics may exhibit qualitatively distinct regimes, for example

$$M(f_t) = 0 \quad \longrightarrow \quad M(f_t) > 0$$

as λ crosses a critical value λ_c. Below λ_c, agents may behave largely independently, while above λ_c, interactions may induce *persistent coordination* or *specialization.* The transition need not correspond to a conventional thermodynamic phase transition; rather, the statistical-mechanical analogy provides a language for identifying sharp changes in the structure of a learning population.

This viewpoint suggests a program for developing **emergence theory** for MARL. Rather than characterizing a learned population only through task performance, we seek to identify order parameters and critical conditions under which qualitatively new collective structures appear. In particular, one can ask i) whether properties such as coordination and specialization are stable, ii) whether multiple collective regimes can coexist, and iii) how these regimes depend on the microscopic interaction rules.

Another important consequence of our proposed framework is that emergence may be **controllable**. To see this, let u_t denote a population-level control signal. The kinetic equation can be conceptually extended to:

$$\partial_t f_t + \mathcal{L}_{\mathrm{MDP}}(V_{f_t}, f_t) = Q(f_t, f_t; u_t).$$

The control u_t may modify *communication, interaction rates, information sharing, incentives, or other aspects of the population dynamics.* The resulting problem is not simply to optimize the policy of an individual agent, but to steer the population distribution toward a desired collective regime:

$$f_0 \xrightarrow{u_t} f_T \in \mathcal{F}_{\mathrm{desired}}.$$

This provides a conceptual foundation for *controlled emergence*: the design of microscopic interaction mechanisms that induce desirable macroscopic organization.

From Emergence to Meta-Cognition

A particularly ambitious form of collective organization is *meta-cognition*: the ability of a learning system to acquire information about its own cognitive state and use that information to regulate subsequent learning and decision making. Our framework suggests that meta-cognition need not be restricted to an individual agent. It may instead emerge as a population-level phenomenon. Toward this end, let $C_t = \mathcal{C}(f_t)$ denote a macroscopic descriptor of the collective cognitive state. Depending on the setting, C_t could summarize *uncertainty, collective competence, coordination, diversity, disagreement, or the effectiveness of information exchange.* The essential feature is that C_t is computed from the population state rather than being explicitly represented by every individual agent.

If the collective state feeds back into the interaction dynamics, we obtain the conceptual feedback loop:

$$f_t \longrightarrow C_t \longrightarrow Q_{C_t} \longrightarrow f_{t+1}.$$

The population therefore does not merely evolve according to fixed interaction rules. Its current collective state can influence how its members interact and learn in the future. For example, a population exhibiting high uncertainty might increase information sharing, whereas a population with high redundancy might favor specialization or exploration. In this sense, meta-cognition can be interpreted as a **self-referential population dynamics**: the population develops a representation of its own state and uses that representation to modify its subsequent evolution.

This perspective also clarifies why the kinetic level is important: An individual agent may have only local observations and limited knowledge of the population, while the distribution f_t contains aggregate information about the entire population. The functional $\mathcal{C}(f_t)$ therefore provides a potential bridge between distributed local learning and population-level self-awareness. Crucially, this does not require assuming that every agent possesses explicit global meta-cognition. Collective meta-cognition may instead arise from the interaction of many locally adaptive agents.

A Dynamical Theory of Collective Intelligence

The combination of emergence and meta-cognition suggests a broader objective: to understand collective intelligence as a dynamical system with identifiable states, transitions, and control mechanisms. At the microscopic level, individual agents optimize their behavior through Bellman dynamics. At the mesoscopic level, interactions reshape the distribution f_t. At the macroscopic level, order parameters such as $M[f_t]$ and collective cognitive states such as C_t characterize the organization of the population.

This hierarchy leads to a new class of questions that are difficult to pose within conventional MARL: 1) Under what conditions does a population spontaneously develop specialized roles? 2) When does communication produce coordination rather than redundant behavior? 3) Can a population transition between qualitatively different collective regimes? 4) Can the onset of such transitions be predicted from microscopic interaction parameters? Most importantly, 5) can we design interactions that reliably induce a desired collective capability?

The ultimate objective is therefore not simply to *observe* emergence, but to develop a theory under which emergence can be *predicted, explained, and controlled*. Within this framework, collective meta-cognition represents an especially strong form of emergence. It corresponds to a population-level feedback loop in which the collective state influences the mechanisms that govern future learning. Establishing rigorous conditions under which such feedback produces stable and useful self-regulation is a major open challenge. Nevertheless, the Bellman-Boltzmann formulation provides a natural starting point because it simultane-

ously represents individual optimality, population evolution, and interaction-mediated feedback. The next section turns to the theoretical foundations required to make this vision mathematically rigorous.

6 Mathematical Foundations

A scientific theory requires more than a plausible model: it requires mathematical results that connect the model to predictions. For the Bellman–Boltzmann framework, we envision a hierarchy of *five* fundamental problems that together connect microscopic multi-agent reinforcement learning to macroscopic collective intelligence.

Problem I: Well-Posedness. The first question is whether a Bellman-Boltzmann system defines a valid population dynamics. For a Bellman-Boltzmann system

$$V_f = \mathcal{T}_f V_f, \qquad \partial_t f + \mathcal{L}_{\mathrm{MDP}} f = Q(f, f),$$

we seek conditions guaranteeing existence and uniqueness of solutions, as well as the fundamental properties:

$$f(s, z, t) \geq 0, \qquad \int f(s, z, t)\, ds\, dz = 1,$$

which ensure that f remains a valid probability distribution throughout its evolution. Kinetic PDE theory, semigroup methods, and Wasserstein analysis provide natural starting points for establishing these properties. This problem is foundational: without well-posedness, neither the subsequent kinetic theory nor predictions of collective behavior are mathematically meaningful.

Problem II: From Agents to Kinetic Theory. The second question is whether the Bellman-Boltzmann equation can be derived from a finite population of interacting learning agents. Given microscopic states $X_1(t), \ldots, X_N(t)$, define the empirical distribution $f_N(t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i(t)}$. Then, we seek to establish convergence $f_N \to f$ as $N \to \infty$, where f satisfies the kinetic equation. A central component is *propagation of chaos*: for every fixed k, we want to show

$$\mathcal{L}(X_1, \ldots, X_k) \longrightarrow f^{\otimes k}.$$

For intelligent agents, this question is substantially richer than in classical particle systems because the microscopic states include adaptive policies and internal learning states. Establishing this limit would provide a rigorous foundation for viewing the kinetic equation as the large-scale limit of MARL rather than merely as a heuristic approximation.

Problem III: A Learning H-Theorem. The third question is whether learning populations possess a fundamental directionality analogous to the Boltzmann H-theorem. Motivated by $H(f) = \int f \log f$, we envision a generalized learning entropy of the form:

$$H_{\mathrm{AI}}(f) = \int f \log f + \lambda C(f),$$

where $C(f)$ captures an appropriate learning or cognitive contribution. The central question is whether one can identify conditions under which

$$\frac{d}{dt} H_{\mathrm{AI}}(f_t) \leq 0.$$

Such a theorem would do more than establish convergence: it could reveal a **universal directional principle** governing learning and self-organization in large populations. Determining the correct functional $C(f)$ is therefore a central theoretical challenge.

Problem IV: Intelligence Emergence as a Bifurcation. The fourth question is whether emergence of collective intelligence can be characterized and induced as a mathematically identifiable *instability* of a population equilibrium. Let f_0 denote a symmetric equilibrium and consider a perturbation $f = f_0 + \epsilon g$. Linearization of the kinetic dynamics yields:

$$\partial_t g = \mathcal{A} g.$$

We hypothesize that the onset of organized collective behavior can be characterized by the spectrum of $\mathcal{A}$, with instability occurring when

$$\lambda_{\max}(\mathcal{A}) > 0.$$

This provides a precise interpretation of emergence as a transition from a disordered regime to an organized regime. It also makes it possible to identify critical interaction parameters and predict when collective structures should appear.

Problem V: Emergent Meta-Cognition. The fifth question is whether meta-cognition (i.e., improving learning itself) can emerge as a macroscopic property of a population of interacting learning agents. Rather than treating meta-cognition as an independent and explicitly engineered module within each individual agent, we envision a population-level cognitive field $m(x, t)$, which represents *collective memory, uncertainty awareness, self-monitoring, and adaptive reasoning* to induce the emergence of meta-cognition. A central challenge is to understand how such a *slow cognitive field* emerges from the faster dynamics of learning and decision making. We therefore envision a multi-timescale kinetic model of the form $\partial_t f = Q_{\mathrm{fast}}(f) + \epsilon Q_{\mathrm{slow}}(f, m)$, $0 < \epsilon \ll 1$, where f captures fast population dynamics while m evolves on a slower cognitive timescale. In meta-cognition, we seek to determine whether the emergence of m can be deliberately controlled through a population-level intervention u (e.g., communication, attention, exploration, and memory). This leads to a meta-cognition control problem: choosing u to minimize intervention cost, subject to the constraint that the desired collective intelligence emerges:

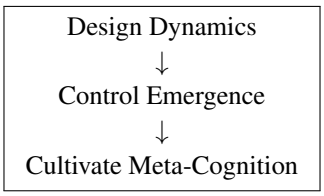
$$\min_{u} \text{ Intervention Costs}$$
$$\text{s.t. } M(f) \geq M_{\mathrm{th}}.$$

where M_{th} is some minimum threshold for intelligence emergence. This problem provides a mathematical foundation for viewing meta-cognition not as a fixed architectural component, but as an emergent thermodynamic variable.

7 Toward a Science of Collective Intelligence

Kinetic Learning Algorithms. The proposed Bellman-Boltzmann framework points toward a broader shift in the paradigm of artificial intelligence. Traditional AI largely follows the pipeline of designing a model, optimizing an objective, and obtaining intelligent behavior. For large populations of agents, however, intelligence may depend as much

on the population dynamics as on the individual agent's capabilities. We envision a complementary paradigm:



Rather than optimizing agents independently, the goal is to shape how their collective state evolves. This perspective suggests a new class of algorithms in which the object is the *population dynamics*. Conventional learning updates individual parameters according to rules such as $\theta_i \leftarrow \theta_i - \eta \nabla L_i$. In contrast, a kinetic learning algorithm would directly manipulate the evolution of the population distribution:

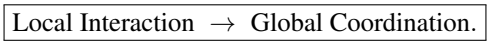
$$f_{t+1} = \mathcal{K}(f_t, u_t),$$

where u_t represents controllable aspects of agent interactions, information exchange, incentives, or learning dynamics. The objective is thus to *shape the distribution of intelligence*: to induce desirable collective structures while preserving the adaptability and heterogeneity of each agent.

Application I: Large-Scale MARL. The first application is large-scale MARL, where AI systems may involve millions of agents with heterogeneous capabilities, decentralized decision making, and dynamic cooperation. At this scale, independently training agents becomes increasingly inadequate for understanding population behavior. The fundamental question changes from *How do we train agents?* to *How do we create intelligent societies?* The Bellman-Boltzmann framework provides a potential mathematical foundation for answering this question by connecting individual learning rules to the evolution of the agent population.

Application II: LLM Agent Societies. Foundation models create a particularly compelling new setting for collective intelligence. Interacting LLM agents can form scientific discovery teams, autonomous organizations, multi-agent reasoning systems, and human–AI collaborative systems. Their collective capabilities may depend not only on the intelligence of individual models but also on communication, role assignment, specialization, information sharing, and feedback among agents. This motivates the development of a mathematical theory of *agent societies* in which emergent capabilities can be analyzed at the population level rather than attributed solely to individual foundation models.

Application III: Embodied Collective Intelligence. The same principles arise in physical multi-agent systems, including robot swarms, autonomous vehicles, and distributed edge-intelligence systems. These systems must transform local observations and interactions into global coordination:



Since physical agents are heterogeneous, distributed, and subject to communication and resource constraints, they provide important testbeds for studying how microscopic interaction rules produce robust macroscopic organization.

Experimental Validation and Scientific Benchmarks. A theory of collective intelligence requires benchmarks that test more than final task performance. We envision experimental protocols for predicting emergence, detecting phase transitions, inducing controlled specialization, measuring communication efficiency, and evaluating meta-cognitive adaptation. A particularly stringent test is whether a theory can predict the onset of a collective transition *before* it is observed experimentally. Such benchmarks would connect the mathematical quantities developed in Section 6—including entropy functionals, stability conditions, bifurcation indicators, and macroscopic order parameters—to observable behavior in large-scale agent populations.

A Research Roadmap. We envision *five stages* for developing this program. Stage I establishes the Bellman-Boltzmann formulation and identifies appropriate microscopic interaction models. Stage II develops the mathematical foundations described in Section 6, including well-posedness, propagation of chaos, a learning H-theorem, emergence as bifurcation, and hydrodynamic limits. Stage III develops kinetic learning algorithms capable of shaping population dynamics. Stage IV validates these algorithms in large-scale AI systems, including MARL, LLM agent societies, and embodied multi-agent systems. Stage V seeks to consolidate these results into a general theory of collective intelligence. The ultimate vision is a transition from *artificial intelligence* to *artificial collective intelligence*. Such a theory should explain not only how individual agents learn, but how intelligence emerges from their interactions, how collective intelligence evolves across scales, and how its emergence can be deliberately controlled. In this sense, the next frontier of AI is not learning alone, but the scientific understanding of *emergence* and *meta-cognition*.

8 Conclusion

In this paper, we proposed a vision for a Bellman–Boltzmann theory of collective intelligence that connects microscopic reinforcement learning, kinetic population dynamics, and macroscopic emergence. The central premise is that large populations of learning agents should be studied not only as collections of individually optimizing policies, but as dynamical systems in which collective intelligence can emerge through interaction. We outlined a mathematical program spanning well-posedness, propagation of chaos, a learning H-theorem, emergence as bifurcation, and hydrodynamic limits, together with new algorithmic and experimental directions for large-scale MARL, LLM agent societies, and embodied collective intelligence. If successful, this program could move the study of multi-agent intelligence from empirical observation toward a predictive science in which emergence is understood, anticipated, and ultimately engineered.

Acknowledgments

This work is supported in part by DARPA YFA D24AP00265, DARPA grant HR0011261E001, NSF grants CAREER CNS2110259, CNS-2112471, ECCS-233104, and ONR N00014-24-1-2729.

References

Ambrose, D. M.; Griffin-Pickering, M.; and Mészáros, A. R. 2024. Kinetic-type Mean Field Games with Non-separable Local Hamiltonians. arXiv:2403.12829.

Boltzmann, L. 1872. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften*, 66: 275–370.

Brinkman, D.; and Ringhofer, C. 2017. A Kinetic Games Framework for Insurance Plans. *Kinetic and Related Models*, 10(1): 93–116.

Burger, M.; Kanzler, L.; and Wolfram, M.-T. 2024. Boltzmann Mean-Field Game Model for Knowledge Growth: Limits to Learning and General Utilities. *Kinetic and Related Models*. To appear / published version.

Burger, M.; Lorz, A.; and Wolfram, M.-T. 2016. On a Boltzmann Mean Field Model for Knowledge Growth. *SIAM Journal on Applied Mathematics*, 76(5): 1799–1818.

Burger, M.; Lorz, A.; and Wolfram, M.-T. 2017. Balanced Growth Path Solutions of a Boltzmann Mean Field Game Model for Knowledge Growth. *Kinetic and Related Models*, 10(1): 117–140.

Guo, X.; Hu, A.; Xu, R.; and Zhang, J. 2019. Learning Mean-Field Games. In *Advances in Neural Information Processing Systems*, volume 32.

Lucas, J., Robert E.; and Moll, B. 2014. Knowledge Growth and the Allocation of Time. *Journal of Political Economy*, 122(1): 1–51.

Saint-Raymond, L. 2009. *Hydrodynamic Limits of the Boltzmann Equation*, volume 1971 of *Lecture Notes in Mathematics*. Berlin, Heidelberg: Springer.