



# Collective cooperative intelligence

Wolfram Barfuss<sup>a,1</sup>, Jessica Flack<sup>b</sup>, Chaitanya S. Gokhale<sup>c,d</sup>, Lewis Hammond<sup>e</sup>, Christian Hilbe<sup>c</sup>, Edward Hughes<sup>f</sup>, Joel Z. Leibo<sup>f</sup>, Tom Lenaerts<sup>g,h,i</sup>, Naomi Leonard<sup>j</sup>, Simon Levin<sup>k</sup>, Udari Madhushani Sehwag<sup>i,2</sup>, Alex McAvoy<sup>m,n</sup>, Janusz M. Meylahn<sup>o</sup>, and Fernando P. Santos<sup>p</sup>

Edited by Marco A. Janssen, Arizona State University, Tempe, AZ; received December 12, 2023; accepted May 6, 2024 by Editorial Board Member Elke U. Weber

Cooperation at scale is critical for achieving a sustainable future for humanity. However, achieving collective, cooperative behavior—in which intelligent actors in complex environments jointly improve their well-being—remains poorly understood. Complex systems science (CSS) provides a rich understanding of collective phenomena, the evolution of cooperation, and the institutions that can sustain both. Yet, much of the theory in this area fails to fully consider individual-level complexity and environmental context—largely for the sake of tractability and because it has not been clear how to do so rigorously. These elements are well captured in multiagent reinforcement learning (MARL), which has recently put focus on cooperative (artificial) intelligence. However, typical MARL simulations can be computationally expensive and challenging to interpret. In this perspective, we propose that bridging CSS and MARL affords new directions forward. Both fields can complement each other in their goals, methods, and scope. MARL offers CSS concrete ways to formalize cognitive processes in dynamic environments. CSS offers MARL improved qualitative insight into emergent collective phenomena. We see this approach as providing the necessary foundations for a proper science of collective, cooperative intelligence. We highlight work that is already heading in this direction and discuss concrete steps for future research.

cooperation | collective action | complex systems science | multiagent reinforcement learning

Cooperation is the ability of a group to successfully and voluntarily act together toward a common interest, even when short-term or individual gains make selfish behavior more appealing to individuals (1). Such situations are commonly referred to as social dilemmas. In a social dilemma, every actor has the incentive to behave selfishly, yet everyone would be better off if all behaved cooperatively. Cooperation is required to maintain environmental commons, local and global, e.g., fisheries, the atmosphere, and biodiversity, but also social commons, such as public infrastructure, education, and healthcare.

The study of whether, when, and how cooperation emerges is an interdisciplinary pursuit involving research fields as diverse as biology, physics, computer science, engineering, and different branches of social sciences. Established mechanisms to achieve and maintain cooperation include outside authorities (2) and bottom-up arrangements based on social reciprocity (3). External authorities can resolve social dilemmas by installing a punishment

or reward scheme, e.g., via taxes and subsidies, that makes selfish actions less attractive to individuals, whereas bottom-up arrangements and social reciprocity find a way to punish defecting behavior through peers (4). However, the challenge of cooperation is far from being solved.

First, large collectives complicate the emergence and robustness of cooperation. Although many mechanisms have been identified that support its emergence and maintenance, it is also widely recognized that effective scaling mechanisms are rare (5): in global public goods, such as the climate, there is no single outside actor with sufficient enforcement power to ensure cooperation authoritatively. In situations involving many, mostly anonymous, participants, reciprocity mechanisms are hard to stabilize (6). Hence, a key challenge for future research remains identifying principles of collective information processing and collective action that offer robust pathways to cooperation in large groups and across scales of organization.

Second, the complexity and variety of human behavior complicate the challenge of cooperation. A more comprehensive understanding of human behavior is required

Author affiliations: <sup>a</sup>Argelander-Chair Integrated System Modeling for Sustainability Transitions, Center for Development Research (ZEF), University of Bonn, Bonn 53113, Germany; <sup>b</sup>Santa Fe Institute, Santa Fe, NM 87501; <sup>c</sup>Max-Planck-Institute for Evolutionary Biology, 24306 Plön, Germany; <sup>d</sup>Center for Computational and Theoretical Biology, Julius-Maximilians University, Würzburg 97074, Germany; <sup>e</sup>Department of Computer Science, University of Oxford, OX1 3QD Oxford, United Kingdom; <sup>f</sup>Google DeepMind, N1C 4DN London, United Kingdom; <sup>g</sup>Machine Learning Group, Département d'Informatique, Université Libre de Bruxelles, 1050 Bruxelles, Belgium; <sup>h</sup>Artificial Intelligence Lab, Computer Science department, Vrije Universiteit Brussel, 1050 Brussels, Belgium; <sup>i</sup>Center for Human Compatible AI, University of California, Berkeley, CA 94720; <sup>j</sup>Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544; <sup>k</sup>Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544; <sup>l</sup>Department of Computer Science, Stanford University, Stanford, CA 94305; <sup>m</sup>School of Data Science and Society, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; <sup>n</sup>Department of Mathematics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; <sup>o</sup>Department of Applied Mathematics, University of Twente, 7522 NB Enschede, The Netherlands; and <sup>p</sup>Faculty of Science, Informatics Institute, University of Amsterdam, 1098 XH Amsterdam, The Netherlands

U.M.S. is an organizer of this Special Feature.

Author contributions: W.B. designed research; W.B. performed research; W.B. analyzed data; and W.B., J.F., C.S.G., L.H., C.H., E.H., J.Z.L., T.L., N.L., S.L., U.M.S., A.M., J.M.M., and F.P.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. M.A.J. is a guest editor invited by the Editorial Board.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](#).

<sup>1</sup>To whom correspondence may be addressed. Email: wbarfuss@uni-bonn.de.

<sup>2</sup>Present address: J. P. Morgan AI Research, New York, NY 10017.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2319948121/-DCSupplemental>.

Published June 16, 2025.

because of its importance in shaping future pathways to sustainability (7). Securing a livable planet and the well-being of future generations requires changes in policy, technology, and, not least, individual behavior. Human decision-making must be conceptualized as a multifaceted cognitive process coevolving with local and global contexts (8). Furthermore, situations involving diverse human actors with heterogeneous needs, preferences, and characteristics are critical to sustainable futures and present a significant challenge to cooperation (9). It is unknown how the diversity among actors and the cognitive complexity within them can offer robust principles for cooperation.

Third, the complexity of the environments in which the actors are embedded makes stabilizing cooperation challenging (10). Challenges arise from 1) feedback (both gradual and abrupt) between the environment and actors' choices; 2) delayed or sparse consequences of actions; and 3) multiple kinds of risks and uncertainties in the environment (11). It is unresolved which factors of dynamic environmental contexts offer robust pathways toward cooperation.

Fourth and last, transient dynamics complicate the challenge of cooperation. While many cooperation-promoting mechanisms have been found, little attention has been paid to transient phenomena toward cooperation (12). This includes time scales on which cooperation is achievable, the critical transition points at which cooperation becomes viable, and the stability and resilience of a cooperative arrangement. All these elements are crucial to consider for sustainability transitions.

To address these challenges, mathematical models are essential. Process-based, mechanistic models allow for theory building and *in silico* experimentation when experimental approaches are either too costly or infeasible (13). The complex systems science (CSS) approach has produced a robust understanding of how interactions among simple identical agents can create macrolevel structures and properties. Notable fields include swarm intelligence, evolutionary games, and population dynamics. However, the microscale individual level in biological, social, and artificial systems consists of sophisticated entities with behaviors and interactions that may not be well-characterized by simplistic representations (14). It is an open question of what collective behavior emerges when interacting entities are capable of sophisticated cognition and embedded in an environmental context.

Agent-based modeling and the field of artificial life allow for complex individual decision rules and agent heterogeneity (15). Yet, this additional complexity often means that the models cannot be analyzed analytically and must be simulated. In addition, many agent-based models suffer from the well-documented criticism of "garbage-in-garbage-out" because the models are easily overparameterized, and the rules afforded to the agents are not necessarily principled nor empirically grounded (16). Multiagent reinforcement learning (MARL) can be viewed as a type of agent-based modeling in which agent behavior does not have to be fixed in advance with plausible heuristics (17). Instead, the agents themselves learn how to behave. However, MARL simulations are usually highly stochastic and computationally expensive, and the often large number of free parameters can make them challenging to interpret (18).

In this paper, we argue that bridging the communities of CSS and MARL can help with these challenges. MARL operationalizes complex individual cognition in dynamic environments. Conversely, CSS enables a principled understanding of the emergence of cooperation, bringing rigorous ideas about emergence and dynamics to MARL for improved interpretation. This is how we understand Collective Cooperative Intelligence. Intelligence refers to the general ability (mainly of a single actor) to achieve a diverse set of goals (19), which is formalized in AI research. Collective intelligence is the intelligence of a group of primarily simple individuals, as primarily studied in CSS (20). Cooperative intelligence is the ability to achieve cooperation in a wide range of contexts (21). Thus, Collective Cooperative Intelligence is the ability of a collective—composed of individuals capable of intelligent decision-making—to act together to improve their common welfare by solving or identifying problems posed by the environment. With AI rapidly advancing, it is imperative to bring these perspectives together (22). Creating tractable models of collective reinforcement learning is crucial for a better understanding of the fundamental principles that drive the emergence of cooperation among complex agents in dynamic environments. Such an understanding can help researchers design more cooperative algorithms and environments and identify critical leverage points toward a sustainable future.

## Background

Here, we juxtapose CSS and MARL approaches to the challenge of cooperation. Their relative differences are summarized in Table 1 while *SI Appendix* provides more details.

**CSS.** Complex systems are generally out of equilibrium with many interacting components, feedback, and couplings between components and levels (23). CSS has been instrumental in studying cooperation through formal models (24). The primary goal of CSS is to understand how macrolevel cooperation emerges from and interacts with simple microlevel processes, before intervening to improve cooperation.

Typically, models aim to explain emerging cooperation from simple yet plausible mechanisms where "cooperation"

**Table 1. Comparison between CSS and MARL approaches to the study of cooperation problems**

	CSS	MARL
Goal	First understand cooperation, then improve it	First improve cooperation, then understand it
Scope	1) Low-dimensional environment 2) Cooperative behavior available	1) High-dimensional environment 2) Cooperative behavior to be learned
Typical performance criteria	1) Level of cooperation 2) Mechanism plausibility 3) Mechanism simplicity	1) Total social welfare 2) Algorithm scalability 3) Generalization
Methodological tools	1) Dynamics 2) Analytics	1) Algorithm design 2) Simulations
Comparative advantages	1) Analytically explainable 2) Computationally lightweight	1) Agent heterogeneity 2) Applicable to large-scale environments

is based on an atomic action with the property that more cooperation translates to higher social welfare (25). For example, the famous strategy tit-for-tat, which merely reciprocates what the opponent did in the previous turn, is surprisingly successful against much more complicated strategies (26). CSS methods are diverse. Typical methods used are evolutionary game theory (27), nonlinear dynamics (28), and complex social network models (29). With its academic origins in theoretical physics and mathematical biology (30), the CSS approach to study cooperation is rooted in dynamics and often captures the analytic conditions for cooperation to emerge (31), which provides a thorough and robust understanding of the cooperation-promoting mechanisms at play. This is facilitated by making the models as simple as possible, which yields low-dimensional models compared to typical MARL simulations.

However, much of the existing theory in this field neglects the complexity of individuals and environmental contexts (14, 32). Humans are capable of foresight, have a theory of mind, make inferences about their environment, and can adapt their behavior correspondingly. There is a clear interest in adapting CSS models to accommodate these nuances (33–36). Considering how such models could be informed by MARL has great potential to unleash novel ways of modeling complex systems to tackle the challenges of collective cooperation in more complex settings.

**MARL.** In a typical MARL setting, each agent observes (parts of) the current state of the environment, then takes an action, after which they observe (part of) the new state of the environment and are provided with a reward indicating how desirable the previous “state-action-state” transition was. Over time, the agents update their strategies (a mapping from observation histories to probability distributions over their action space) to optimize the long-term amount of reward they receive (37). Studies of cooperation in MARL fall under the umbrella of Cooperative AI (21) with the overarching aim to improve the cooperative capabilities of advanced AI systems by prescribing how agents should (learn to) act. Extending machine learning interpretability techniques to MARL is an ongoing effort to advance also the understanding of MARL systems (38).

In contrast to CSS studies, cooperation is typically not readily available as an action. Instead, what constitutes a cooperative strategy, and how such a behavior can be implemented, must be learned from scratch (39) based on the performance criterion of total social welfare. Furthermore, the learning algorithms developed in MARL should generalize to novel situations and scale to high-dimensional environments. Modern MARL is inspired by a number of fields, including neuroscience, psychology, economics, and machine learning (40–43). For example, the commonly used idea of temporal-difference learning is based upon reward-prediction errors, common to humans, other animals, and machines. In recent years, these traditional ideas have been combined with advances in deep learning, achieving remarkable successes (18). Thus, the methodological focus often lies in designing novel algorithmic features to improve the cooperativeness of RL algorithms in large-scale computer simulations.

However, MARL simulations—on their own—do not facilitate analytically reliable insights into how collective cooperation emerges from complex human and machine behavior in dynamic environments. They often require significant computational resources, while the space to explore suffers from the curse of dimensionality. Moreover, they are typically highly stochastic in nature, and results can be difficult to interpret (18). We believe that a unified approach that combines approaches from CSS and MARL could fill this gap.

## Building Bridges

Here, we discuss how CSS and MARL can complement each other. First, we establish a shared language by treating MARL as a complex dynamical system. Second, we illustrate how CSS can offer MARL improved qualitative insight into emergent collective learning dynamics via its mathematical toolbox. Third, we discuss the potential of MARL for improving CSS approaches regarding formalizing collective behavior from cognitive processes in dynamic environments. We highlight steps for future research throughout the discussion, which we summarize in Box 3.

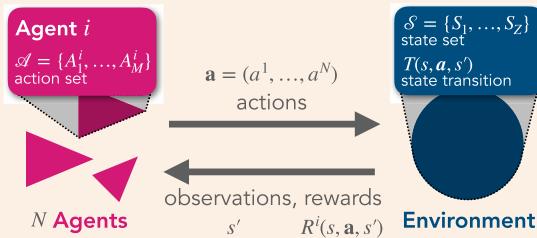
**Collective Reinforcement Learning Dynamics (CRLD).** Using nonlinear dynamical systems to model MARL is an interdisciplinary pursuit involving perspectives from economics (42, 45), sociology (46, 47) machine learning (48, 49), control theory and engineering (50, 51), statistical mechanics (52, 53), and mathematical biology (54, 55). These fields have diverse research goals. Economics focuses on convergence to equilibrium solutions, whereas statistical mechanics is interested in nonconvergence scenarios. Machine learning and engineering care for the development of scalable algorithms and evaluation metrics. Mathematical biology and sociology view reinforcement learning as a model of biologically plausible adaptation and study when and how certain learning processes serve as a model of human behavior. To overcome disciplinary divides, we summarize the intersection of these approaches under the term CRLD and present an exemplary illustration in Box 1.

However, the study of cooperation has not been at the center of studies despite notable examples existing in mathematical biology and sociology (see *SI Appendix*). Here, we argue for putting the question of cooperation at the center of a research agenda on collective cooperative intelligence, accompanied by regarding learning as a biologically plausible model of human and AI behavior and focusing on the complex, transient learning dynamics. See Box 2 for a minimal application example.

Methodologically, CRLD studies are characterized by two forms of idealization. First, CRLD focuses on low-dimensional environments compared to typical MARL studies. Often, just static games with two agents are investigated. Second, CRLD simplifies the highly stochastic and computationally intense reinforcement learning algorithms into deterministic (differential or difference) learning equations. The connection between learning algorithms and equations either stems from stochastic approximation theory (51), or evolutionary game theory (28). The replicator dynamics from mathematical biology describe not only evolutionary

**Box 1.****Collective Reinforcement Learning Dynamics.**

CRLD provides a bridge between CSS and MARL based on the mathematical framework of MARL:



At each point in time, each agent  $i \in \{1, \dots, N\}$  can choose an action from its action set  $A^i$ . These agents are embedded in a (physical, ecological, or social) environment with possible states  $\mathcal{S}$ . States change according to the environmental transition function: a probability of entering the next state  $s'$ , for each state  $s$  and each combination of agents' actions  $\mathbf{a}$ . Agents receive external rewards: a numeric value for each agent  $i$ , for each transition from  $s$  via action  $\mathbf{a}$  to state  $s'$ . Instead of RL algorithms, CRLD uses a set of dynamical equations, modeling the learning of the agents. As with algorithms, many variants exist. Here, we illustrate the dynamic equations of temporal-difference learning (44):

$$X_{t+1}^i(s, a) = \frac{1}{3} X_t^i(s, a) \exp(\eta^i \cdot \delta_{X_t}^i(s, a))$$

The joint strategy, i.e., the probability of each agent  $i$  choosing action  $a$  in state  $s$ ,  $X_{t+1}^i(s, a)$ , is updated by the product of the previous strategy,  $X_t^i(s, a)$ , and the exponent of the previous strategy-average temporal-difference error,  $\delta_{X_t}^i(s, a)$ , multiplied with the effective learning rate  $\eta^i$ , and properly normalized,  $1/3 X_t^i(s)$ , to yield a probability distribution.  $\delta_{X_t}^i(s, a)$  tells the agents how to adapt their strategy to gain more reward over time, averaging out all randomness due to the stochasticity of agents' strategies and environmental transitions. See *SI Appendix* for all details. The benefits of this framework are described below for MARL in Complex Phenomena and for CSS in Cognition in Contexts.

individual learning schemes arise, how they are related, and which algorithmic details matter on the collective level. Such a CRLD theory would enable us to summarize the idealized learning behavior of a broad set of algorithms. See refs. 51, 59, and 60 for promising starting points.

**Complex Phenomena.** Statistical mechanics provides a formal framework to derive emergent phenomena from microscopic rules for systems that are too complex to analyze directly (61). The critical advantage of CRLD is the ability to uncover complex emergent phenomena within a MARL system in a computationally fast and lightweight manner. In this section, we highlight how a CSS perspective enriches the study of cooperation among intelligent agents in complex environments in MARL, using the framework outlined in Box 1 and paving the way toward a "strategic statistical mechanics" (62).

**Multistability.** MARL studies typically report learning success over time. CSS aims to analyze the whole dynamics emerging from agents learning in the environment. This can be done by visualizing (projections of) the joint strategy space. For example, Fig. 1A shows the phase space in the prosperous state for the environmental example of Box 2 with the exemplary learning dynamics of Box 1. Each arrow represents the average effective reward-prediction error the agents perceive at this strategy point. They, therefore, give an intuition about the direction in which the collective will learn.

A phase space perspective highlights the stability landscape, e.g., indicating whether the current situation exhibits more than one equilibrium. For example, the blue learning trajectories in Fig. 1A enter the mutual-cooperation strategy equilibrium point. The red learning trajectories go to mutual defection. Assuming an equal likelihood of initial strategies, the size of the so-called basin of attraction for a given equilibrium indicates how likely the equilibrium will occur.

Multistability in collective learning has exciting applications. First, the size of the basin of attraction of the mutually cooperative solution is a valuable measure of collective (cooperative) intelligence (63). Second, stability landscapes and basins of attraction are instrumental to the study of social–ecological resilience (64), which make CRLD an ideal candidate to advance this field (65). Third, entering one over other possible equilibria is a form of storing information and thus can be seen as a form of emergent collective memory (66). This perspective can explain social conventions like technical standards, cultural norms, and moral rules. Applying it to engineering MARL systems offers the potential to outsource cognitive functions onto the collective. Fourth and last, the relative size of attraction basins offers a valuable perspective for ad hoc teamwork (67). It provides a geometric view of how new agents impact equilibria selection in groups of agents trained a priori when new agents—possibly with strategies hardcoded in a particular way (68)—are introduced in the system.

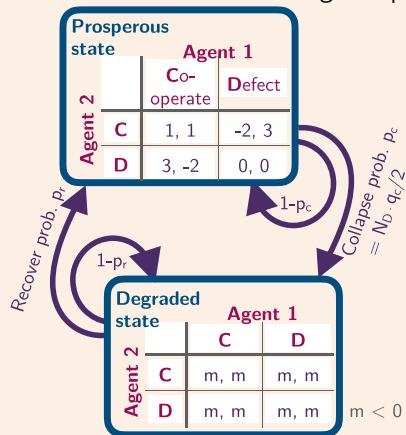
Around the basin of attraction's edge, also known as "separatrix" (dashed purple line in Fig. 1A), the transient dynamics before entering equilibria show complex phenomena. For example, the fine-grained bundle of learning trajectories shows that, at the separatrix, only a tiny differ-

processes but also individual learning processes (56). The relationship between the two fields is as follows: one population with a distribution over phenotypes in the evolutionary setting corresponds to one agent with a distribution over actions in the learning setting (57). Crucially, unlike other dynamical system methods that describe the macroscopic behavior of many agents in a low-dimensional dynamical system, CRLD can also describe the idealized learning behavior (58) of a single or few reinforcement learning agents.

What is still missing is an integrated theory of CRLD (Box 3), which explains from what principles different

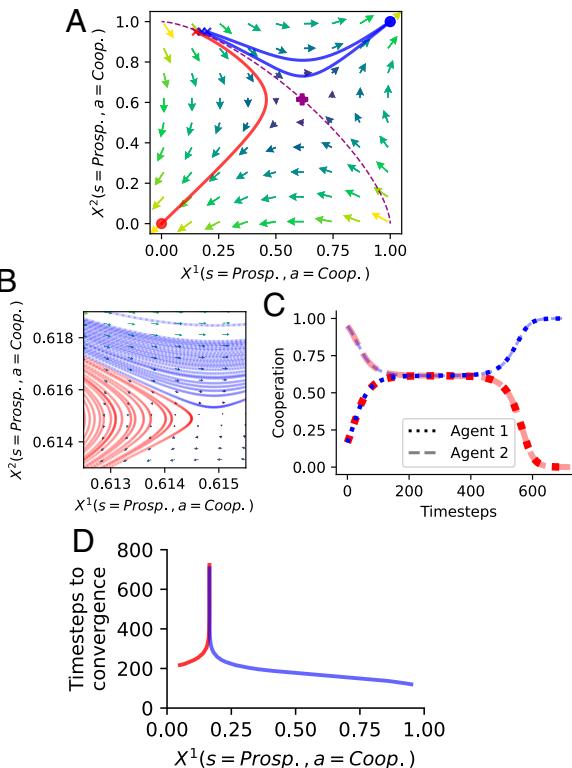
**Box 2.****Example environment.**

Here, we illustrate an environment that allows for the study of collective action under ecological tipping.

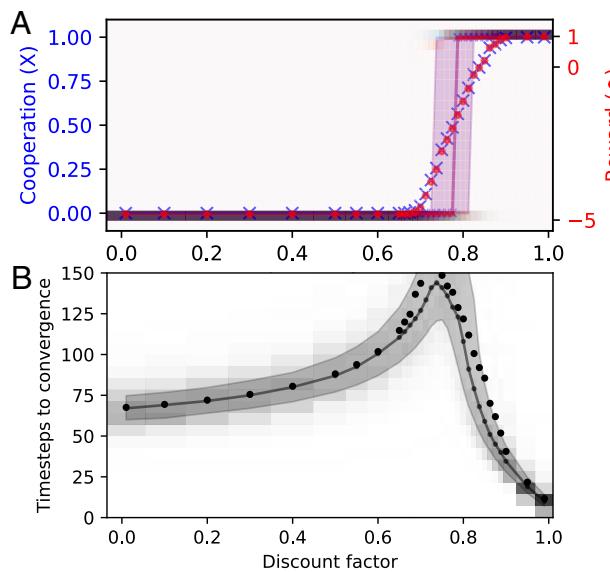


Two agents play a standard public goods game. Everyone is better off by cooperating but each agent has an immediate incentive to exploit the other. However, here, the agents are embedded in a dynamic environment consisting of two states, a prosperous and a degraded one. Each defecting agent increases the probability of collapse by an amount  $q_c/2$ . When collapsing to the degraded state, each agent suffers from the collapse by a negative impact  $m$  at each time point until they recover to the prosperous state, governed by the recovery probability  $p_r$ . For Figs. 1–3, we use  $m = -5$ ,  $q_c = 0.2$ ,  $q_r = 0.01$ .

**Critical transitions.** When small changes in external parameters have significant consequences for the system's behavior, CSS speaks of a critical transition. Other terms for this phenomenon are regime shifts, bifurcations, tipping elements, or phase transitions. Some CRLD studies have already investigated such critical transitions in collective learning (53, 55, 69–72). External parameters in the MARL framework consist of the learning hyperparameters and the parameters defining the environment. Hyperparameters denote the parameters of the learning process, as opposed to parameters used to encode an agent's strategy. An example of an agent's hyperparameter is its learning rate, which determines the extent newly obtained information overrides old information. Parameters defining the environment shown in Box 2 are the severity and likelihood of collapse,  $m$ , and  $q_c$ . The analogy to phase transitions from classical statistical mechanics is as follows: a system (such as water) can exist in multiple stable forms (liquid, ice, steam) depending on external parameters (such as pressure and temperature). For standard room pressure, the transitions between these stable forms happen around the critical temperature points of 0 and 100 °C. Fig. 2A shows the transition from complete defection to full cooperation by varying the agents' discount factor, denoting how much the agents care for their future wellbeing. For discount factors below 0.7, full defection is globally stable (i.e., reached for any initial condition), and for those above 0.85, full cooperation is globally stable. In between, they are in a bistable regime where both full cooperation and full defection are stable



**Fig. 1.** Multistability in CRLD (Box 1) applied to the ecological tipping environment (Box 2). (A) Phase space of the prosperous state. (B) Detailed bundle of learning trajectories. (C) Emergent timescales with abrupt transitions in transient dynamics. (D) Critical slowing down around the critical point. See *SI Appendix* for details.



**Fig. 2.** Critical transitions in CRLD (Box 1) applied to the ecological tipping environment (Box 2). (A) Cooperation levels and final rewards. (B) Time steps to convergence show a critical slowing down around the critical point. For each discount factor, both plots show a histogram of converged results from 250 random initial conditions by a color map, their mean by large markers, their median by small markers, and the range between lower and upper quartiles by the shaded regions. See *SI Appendix* for details.

(as in Fig. 1A). This can be seen by the abrupt change in the quartiles in Fig. 2A, whereas the smooth change in the average indicates a continuously shifting basin of attraction from full defection to full cooperation.

Near the critical region, the phenomenon of critical slowing down can be observed again. For discount factors of around 0.75, learning takes approximately twice as long as for small discount factors and about an order of magnitude longer than for high discount factors (Fig. 2B). While this phenomenon emphasizes the need to focus more on the transient learning dynamics, it also offers a potential application. So-called “early warning indicators” from CSS (73) could be utilized in collective learning processes to detect and proactively address nearby transition points.

Overall, the occurrence of critical transitions in multiagent learning makes CRLD a promising modeling tool for studying emergent social tipping points in social–ecological systems and human–environment interactions.

**Hysteresis.** If we revert the external parameter after a critical transition, but the system does not follow the same trajectory back, we observe the phenomenon known as “hysteresis” (74–77). Fig. 3 shows hysteresis in CRLD. In contrast to Fig. 2, where the CRLD system was simulated to convergence for each parameter point independently, in Fig. 3, the external parameter (here, the discount factor) changes slowly within the simulation. Around a discount factor of 0.83, the CRLD system tips from defection to cooperation. It remains there until the discount factor decreases again until 0.7 when it tips again back to full defection. For discount factors between 0.7 and 0.83, the agents’ strategy depends on where the discount factor was before. More complex phenomena might arise when the timescales of external parameter changes approach the internal timescale of the learning dynamics (78).

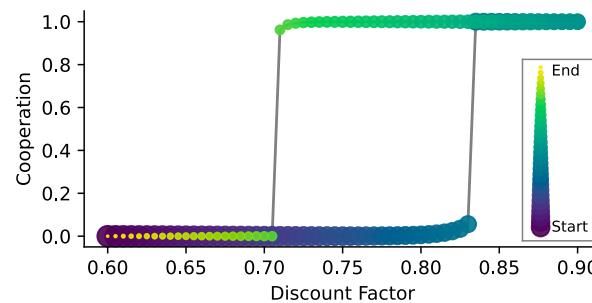
Hysteresis means that the system’s state is dependent on its history of external parameter changes, i.e., it can store information. Thus, hysteresis in MARL presents another form of collective memory. Future work on (large-scale) MARL systems can benefit from more explicitly considering the complex networks between agents to outsource cognitive functions between the agents. For instance, in the line of work on autocurricula, a changing distribution of environments might endow (collectives of) agents with (cooperative) skills that robustly persist even as the distribution of environments changes over the course of learning.

Taken together, the complex emergent phenomena around multistability, critical transitions and hysteresis plus a plurality of possible dynamic regimes as shown in the *SI Appendix* in CSS offer great potential for analyzing and designing MARL systems. These phenomena are observable in CRLD, making it highly likely that they can also be found within high-dimensional MARL systems. However, it is still an open research question to systematically investigate the conditions for the occurrence and the potential use cases for emergent phenomena in MARL (Box 3).

**Cognition in Contexts.** Last, we discuss how ideas from MARL can enrich CSS regarding the three challenge areas of individual cognition, collective behavior, and dynamic environments.

**Cooperation from individual cognition.** First, we explore how cooperation might emerge from individual cognition and what is needed to move CRLD to a position able to investigate this question.

CSS’s dominant interpretation of how successful strategies spread is by agents copying the strategies of more successful agents. This approach has been very influential in capturing the key elements of cultural evolution (79). However, this form of social learning assumes that agents share the same success criteria and can observe both strategies and the success of others, making it difficult to apply to asymmetric interactions. There is clear interest within CSS in extending this paradigm to other forms of cognition and learning (32, 80). The various cognitive mechanisms explored in CSS do not explain how to integrate them into one coherent model. Such integration, however, is necessary to examine their interplay and relative importance for emerging collective cooperation in dynamic environments. Moreover, cognition is general-purpose and must, therefore, be examined across different environments.



**Fig. 3.** Hysteresis in CRLD (Box 1) applied to the ecological tipping environment (Box 2). The discount factor changes while the agents keep on learning. The size and color of the dots represent the time, from dark to light and from big to small as from past to future. See *SI Appendix* for details.

MARL offers a comprehensive framework for studying the interplay among learning, representation, and decision-making between multiple actors (81), offering an integrating platform to test hypotheses on how different cognitive mechanisms may affect collective cooperation in dynamic environments. For example, intrinsic motivations guide learning without relying on externally supplied rewards for improved exploration, control, and homeostasis (82). Being explicitly prosocial and including the rewards of others in one's own reward function (i.e., having other-regarding preferences) clearly promotes cooperation (83). In general, however, it is nonobvious how other common types of intrinsic motivations affect cooperation, e.g., being curious (84), being cautious or risk-taking (85), being controlling (86), or being predictive (87). Even less is known about how other cognitive functions affect collective cooperation in dynamic environments. For example, a key component in single-player deep RL is experience-replay, bearing some resemblance to the replay events observed in the hippocampus (88). Other cognitive mechanisms are employed to learn

### Box 3.

#### Open research strands.

The following open research areas are intended to cross-fertilize progress on collective cooperative intelligence between CSS and MARL:

1. **CRLD Theory.** What are the principles that give rise to different RL update schemes? How are they related? Which algorithmic details matter on the collective level and which do not?
2. **Complex dynamic phenomena.** What are the conditions for the occurrence of and the potential use-cases for complex emergent and transient phenomena, such as multistability, abrupt transitions, hysteresis, and dynamic regimes in CRLD and large-scale MARL?
3. **CRLD with cognition.** How can cognitive mechanisms such as representational learning, world models, intrinsic motivations, and theory of mind be integrated into CRLD? What are their effects on the learning of cooperation? What is the role of intrinsic noise in CRLD?
4. **CRLD in large collectives.** What are the assumptions that result in mean-field approaches to MARL? How do they relate to each other and to CRLD? And what are the principles in CRLD from which cooperation can emerge in large collectives composed of intelligent individuals, particularly heterogeneous ones? How can these principles be transferred to large-scale MARL systems?
5. **CRLD in dynamic environments.** How do environments of different levels of abstraction relate to each other? What techniques are suitable for scaling environments up and down? Which environmental properties promote or hinder the learning of cooperation? How do these properties influence large-scale MARL systems?

efficient representations of high-dimensional observations (89) and internal models of the world (90). These mechanisms are crucial for agents to update their behavior in a robust and efficient way.

The benefit of CRLD is that it combines the integrating platform of classical MARL for testing hypotheses regarding cognition (Box 1) with the analysis toolbox of CSS. Deterministic CRLD model agents learn in the theoretical limit of an infinite memory batch (91) or as if they have a perfect model of the current environment (58). It is such abstractions that make these models useful. By comparing the deterministic CRLD dynamics with the stochastic MARL simulations, it has been shown that the intrinsic stochasticity of the MARL process is highly beneficial for the learning of cooperation (92). Lenient CRLD dynamics make agents forgiving to initial mismatched teammate actions, which result in a higher likelihood of converging to the cooperative solution. CRLD with partial observability can learn to cooperate through inaccurate representations of the environment (93). However, future research is needed to examine the role of more refined notions of cognition, such as adaptive intrinsic representations, world models, motivations, and theory of mind in the learning of cooperation, in general, and of intrinsic noise in the emergence of cooperation in dynamic environments by means of stochastic dynamical systems (94), in particular (Box 3).

**Cooperation in large collectives.** Next, we discuss the challenge of cooperation in large collectives of self-learning agents in dynamic environments and how CRLD can investigate this problem.

In CSS, the canonical approach assumes a large collective of homogeneous and/or simple individuals, and there is a long tradition of describing the phenomena emerging from such collectives (28). For example, the famous replicator equation describes the evolutionary dynamics of a population containing an infinite number of individuals (95) and many works have investigated the preconditions for cooperation under this paradigm. Refining this paradigm transforms evolutionary dynamics from infinite to finite (but still large) populations (96).

In MARL, large collectives are much harder to study because of the exponentially increasing joint state-action space and the associated computational demands that come with it. However, some notable examples exist (97, 98). They manage the computational demands by careful software engineering, exploiting agent similarities through parameter sharing or using transfer learning mechanisms while progressively increasing the number of agents. Yet, little is known about cooperation in large collectives from a MARL perspective.

In CRLD, a few works have begun to explicitly treat a collective of RL agents as a nonlinear dynamical system, deriving a Fokker–Planck equation to describe the idealized learning behavior of a MARL collective. They focus on repeated symmetric games (49), population games (99), stochastic effects and incomplete information (100), regret minimization (101), and regular social network structures (102). There is an apparent similarity here with so-called mean-field approaches to MARL. MARL is often very high-dimensional and highly stochastic, making it computationally challenging and difficult to understand. Mean-field

approaches, originally from theoretical physics, attempt to solve this problem by approximating the original system with a simpler model by averaging over certain degrees of freedom. It is, therefore, an ideal tool for gaining conceptual clarity about collective cooperation in complex settings. However, being a research frontier, mean-field approaches to RL are not yet consolidated, and different variants exist. Most prominently, mean-field games (103) consider the large-agent limit in a situation where each agent has an individual state associated with it. Other works consider the mean action of a local neighborhood of agents (104), which allows the study of a finite number of agents. All variants primarily focus on obtaining convergence guarantees. Future work is needed to distill the emergent phenomena, transient dynamics, and principles that lie behind cooperation in large collectives composed of intelligent individuals, particularly heterogeneous ones from a CRLD perspective. To do so, it will be vital to consolidate these different mean-field approaches, contrast the assumptions from which these arise, and refine them to heterogeneous and structured populations (Box 3).

**Cooperation in dynamic environments.** Last, we explore the challenge of cooperation in dynamic, stateful environments and what is required to leverage the most out of the CRLD approach.

In CSS, recent years have seen a growing interest in moving from evolutionary dynamics in stateless games to dynamic environments. Here, the term “dynamic” can mean external fluctuations (105), a varying population density (106), spatial network structure (107), or coupled systems of evolutionary and environmental dynamics. Coupled systems may be categorized into those with continuous environmental state spaces (108) or discrete ones (109). These examples reflect a diverse understanding of the term “environment.” However, agents typically remain in the paradigm of social learning in these works.

The MARL setting employs large, dynamic, and uncertain environments via the framework of (partially observable) stochastic games. Sequential social dilemmas form a bridge between high-dimensional environments and the simple, stateless games primarily used to study cooperation in CSS (39, 83). They share simple games’ mixed incentive structure but require agents to learn cooperation strategies on their own, i.e., cooperation is not a single atomic choice but a temporally extended sequence of actions. See *SI Appendix* on how to define “cooperation” in stateful environments. For such complex environments, deep MARL can also be used to generate predictions of human behavior, where traditional game-theoretic methods are no longer amenable (110).

Some CRLD works already consider dynamic environments via learning in (partially observable) stochastic games (44, 111). To leverage their strengths in abstraction, CRLD studies should focus on comparably low-dimensional stochastic games with few environmental states to combine the best of both worlds: analytical tractability and environmental complexity. For example, our guiding example (Box 2) is a stochastic game with only two states. Nevertheless, it is sufficient to highlight a range of complex phenomena (see above) and calculate analytically when the level of caring for future rewards can turn the tragedy of the

commons into a comedy—where cooperation prevails—without requiring any form of social reciprocity (91). Future work is required to explore more generally which environmental conditions favor or hinder the learning of cooperation (Box 3). Factors to consider are environmental uncertainty, delayed and sparse rewards, the coupling of local and global environmental states, environmental tipping points, and spatial extendedness. To do so, a better understanding of how environments on different abstraction levels are related. Empirical Game Theoretic Analysis downscals a high-dimensional environment to a stateless game via heuristic strategies (112). Intuitively, this technique complements the sequential social dilemma concept (39), reducing an environmentally complex situation to its core. Future work should aim to downscale a high-dimensional environment not necessarily to a single environmental state but to a few significant states to preserve the essential environmental context (see also *SI Appendix*). For this purpose, conceptual links to temporally extended substrategies of hierarchical reinforcement learning may be fruitful (113).

## Conclusions

Collective cooperation—in which intelligent actors in complex environments seek ways to improve their joint well-being—is critical for a sustainable future, yet unresolved. Mathematical models are essential to solve this challenge. Here, we argue that building bridges between CSS and MARL offers a more robust understanding of the drivers, mechanisms, and dynamics of collective cooperation from intelligent actors in dynamic environments. Both fields complement each other in their goals, methods, and scope.

CSS offers MARL improved qualitative insight into emergent collective learning dynamics via its analysis toolbox. Based on the existing body of works that models MARL as nonlinear dynamical systems, we highlighted what kinds of insights a CSS perspective on complex, emergent phenomena and dynamics offers. In our guiding example (Box 2), all three elements of cognition, collective, and environment are represented in their most minimal form.

By way of return, MARL offers CSS concrete ways to formalize cognitive processes in dynamic environments. We portrayed the current state of cooperation research along the three elements of cognition, large collectives, and dynamic environments. Box 3 summarizes the five open research strands we identified. Ultimately, by sharing the same mathematical framework, MARL and CSS complement each other in their scope. CSS can focus on more qualitative, conceptual models with comparable low-dimensional environments. In contrast, MARL can focus on scaling cooperation-promoting principles to more quantitative, realistic models with comparable high-dimensional environments.

This complementary view of MARL and CSS can also strengthen the role of theory. It enables the formulation of integrated theories around the dynamics of cooperation in large collectives from individual cognition in environmental contexts. The connection between empirical data

is threefold. 1) On the behavioral level, theory predictions can be compared with behavioral experiments or field observations. In addition to procedures with human or nonhuman animal subjects, large language models (LLMs) offer innovative ways to investigate various phenomena in human-machine ecologies. 2) On the neurological level, stylized theoretical representations of neural information processing could be compared with natural or artificial brain data (as obtained, e.g., in neuroeconomics). 3) Biophysical and socioeconomic data can be used to parameterize the model environment. The resulting learning dynamics offer projections of future socioeconomic pathways, and the distance to critical points in parameter space gives a measure of their resilience.

We believe that the approach of building bridges between CSS and MARL provides the necessary foundations for a genuine science of collective, cooperative intelligence and invite everyone to join forces.

**Data, Materials, and Software Availability.** All supplementary information in fully reproducible form is available here (114) or in the [SI Appendix](#).

**ACKNOWLEDGMENTS.** W.B. acknowledges support from the Cooperative AI Foundation. C.S.G. acknowledges support from the Max Planck Society and the Julius-Maximilians University Würzburg. C.H. acknowledges generous funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Starting Grant 850529: E-DIRECT) and from the Max Planck Society. T.L. is supported by an F.W.O. project (Grant No. G054919N) and two F.R.S.-FNRS PDR (Grant Nos. 31257234 and 40007793). T.L. is furthermore supported by Service Public de Wallonie Recherche under Grant No. 2010235-ariac by digitalwallonia4.ai, the Flemish Government through the AI Research Program and TAILOR, a project funded by EU Horizon 2020 research and innovation program under GA No. 952215. F.P.S. acknowledges funding from the European Union (ERC Starting Grant, RE-LINK, 101116987). Disclaimer: This work was performed prior to [U.M.S.] joining J.P. Morgan Chase & Co. and should not be considered a research product of that company.

1. S. Levin, "Collective cooperation: From ecological communities to global governance and back" in *Unsolved Problems in Ecology*, A. Dobson , D. Tilman, R. D. Holt, Eds. (Princeton University Press, 2020), pp. 311-317.
2. G. Hardin, The Tragedy of the Commons: The population problem has no technical solution; it requires a fundamental extension in morality. *Science* **162**, 1243-1248 (1968).
3. E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, 1990).
4. B. Rockenbach, M. Milinski, The efficient interaction of indirect reciprocity and costly punishment. *Nature* **444**, 718-723 (2006).
5. R. Boyd, P. J. Richerson, The evolution of reciprocity in sizeable groups. *J. Theor. Biol.* **132**, 337-356 (1988).
6. S. Bowles, H. Gintis, "Cooperation" in *The New Palgrave Dictionary of Economics*, M. Vervengo, E. Perez-Caldentey, J. Ghosh, Eds. (Palgrave Macmillan, London, UK, 2016), pp. 1-10.
7. C. Schill *et al.*, A more dynamic understanding of human behaviour for the Anthropocene. *Nat. Sustainability* **2**, 1075-1082 (2019).
8. S. M. Constantino, M. Schlüter, E. U. Weber, N. Wijermars, Cognition and behavior in context: A framework and theories to explain natural resource use decisions in social-ecological systems. *Sustainability Sci.* **16**, 1651-1671 (2021).
9. A. Tavoni, A. Dannenberg, G. Kallis, A. Loschel, Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11825-11829 (2011).
10. M. A. Janssen, Introducing ecological dynamics into common-pool resource experiments. *Ecol. Soc.* **15**, 7 (2010).
11. S. Levin, A. Xepapadeas, On the coevolution of economic and ecological systems. *Annu. Rev. Res. Econ.* **13**, 355-377 (2021).
12. D. Andersson, S. Bratsberg, A. K. Ringsmuth, A. S. de Wijn, Dynamics of collective action to conserve a large common-pool resource. *Sci. Rep.* **11**, 9208 (2021).
13. S. Levin, Complex adaptive systems: Exploring the known, the unknown and the unknowable. *Bull. Am. Math. Soc.* **40**, 3-19 (2003).
14. J. Flack, "Life's information hierarchy" in *From Matter to Life: Information and Causality*, G. F. R. Ellis, P. C. W. Davies, S. I. Walker, Eds. (Cambridge University Press, Cambridge, 2017), pp. 283-302.
15. J. M. Epstein, R. L. Axtell, *Growing Artificial Societies: Social Science From the Bottom Up* (Brookings Institution Press, Washington, DC, ed. 1, 1996).
16. J. Schulze, B. Müller, J. Groeneveld, V. Grimm, Agent-based modelling of social-ecological systems: Achievements, challenges, and a way forward. *J. Artif. Soc. Soc. Simul.* **20**, 8 (2017).
17. M. P. Wellman, Putting the agent in agent-based modeling. *Auton. Agents Multi-Agent Syst.* **30**, 1175-1189 (2016).
18. P. Hernandez-Leal, B. Kartal, M. E. Taylor, A survey and critique of multiagent deep reinforcement learning. *Auton. Agents Multi-Agent Syst.* **33**, 750-797 (2019).
19. S. Legg, M. Hutter, Universal intelligence: A definition of machine intelligence. *Mind Mach.* **17**, 391-444 (2007).
20. J. Flack, P. Ipeirotis, T. W. Malone, G. Mulgan, S. E. Page, Editorial to the inaugural issue of collective intelligence. *Collect. Intell.* **1**, 26339137221114179 (2022).
21. A. Dafoe *et al.*, Cooperative AI: Machines must learn to find common ground. *Nature* **593**, 33-36 (2021).
22. I. Rahwan *et al.*, Machine behaviour. *Nature* **568**, 477-486 (2019).
23. S. Levin, Complex adaptive systems: Exploring the known, the unknown and the unknowable. *Bull. Am. Math. Soc.* **40**, 3-19 (2002).
24. R. Axelrod, *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration* (Princeton University Press, 1997).
25. A. McAvoy, Y. Mori, J. B. Plotkin, Selfish optimization and collective learning in populations. *Physica D* **439**, 133426 (2022).
26. R. Axelrod, W. D. Hamilton, The evolution of cooperation. *Science* **211**, 1390-1396 (1981).
27. M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, 2006).
28. J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, ed. 1, 1998).
29. M. Perc, J. Gómez-Gardenes, A. Szolnoki, L. M. Floría, Y. Moreno, Evolutionary dynamics of group interactions on structured populations: A review. *J. R. Soc. Interface* **10**, 20120997 (2013).
30. S. Strogatz *et al.*, Fifty years of 'More is different'. *Nat. Rev. Phys.* **4**, 508-510 (2022).
31. M. A. Nowak, Five rules for the evolution of cooperation. *Science* **314**, 1560-1563 (2006).
32. J. M. McNamara, Towards a richer evolutionary game theory. *J. R. Soc. Interface* **10**, 20130544 (2013).
33. C. Hilbe, Š Šimsa, K. Chatterjee, M. A. Nowak, Evolution of cooperation in stochastic games. *Nature* **559**, 246-249 (2018).
34. O. P. Hauser, C. Hilbe, K. Chatterjee, M. A. Nowak, Social dilemmas among unequal. *Nature* **572**, 524-527 (2019).
35. X. Wang, F. Fu, Eco-evolutionary dynamics with environmental feedback: Cooperation in a changing world. *Europys. Lett.* **132**, 10001 (2020).
36. J. M. McNamara, A. I. Houston, O. Leimar, Learning, exploitation and bias in games. *PLoS ONE* **16**, e0246588 (2021).
37. L. Busoni, R. Babuska, B. De Schutter, A comprehensive survey of multiagent reinforcement learning. *IEEE Trans. Syst. Man, Cybernet. Part C (Appl. Rev.)* **38**, 156-172 (2008).
38. T. McGrath *et al.*, Acquisition of chess knowledge in AlphaZero. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2206625119 (2022).
39. J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, T. Graepel, "Multi-agent reinforcement learning in sequential social dilemmas" in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, AAMAS '17* (2017), pp. 464-473.
40. P. Dayan, Y. Niv, Reinforcement learning: The Good, The Bad and The Ugly. *Curr. Opin. Neurobiol.* **18**, 185-196 (2008).
41. R. R. Bush, F. Mosteller, A mathematical model for simple learning. *Psychol. Rev.* **58**, 313-323 (1951).
42. D. Fudenberg, D. K. Levine, "The theory of learning in games" in *Economic Learning and Social Evolution*, K. Binmore, Ed. (MIT Press, Cambridge, MA, 1998).
43. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction, Adaptive Computation and Machine Learning Series* (The MIT Press, Cambridge, MA, ed. 2, 2018).
44. W. Barfuss, J. F. Donges, J. Kurths, Deterministic limit of temporal difference reinforcement learning for stochastic games. *Phys. Rev. E* **99**, 043305 (2019).
45. I. Erev, A. E. Roth, Maximization, learning, and economic behavior. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10818-10825 (2014).
46. M. W. Macy, A. Flache, Learning dynamics in social dilemmas. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7229-7236 (2002).
47. J. Zschache, Melioration learning in iterated public goods games: The impact of exploratory noise. *J. Math. Sociol.* **42**, 1-16 (2018).
48. D. Bloembergen, K. Tuyls, D. Hennes, M. Kaisers, Evolutionary dynamics of multi-agent learning: A survey. *J. Artif. Intell. Res.* **53**, 659-697 (2015).
49. S. Hu, C. W. Leung, H. F. Leung, "Modelling the dynamics of multiagent Q-learning in repeated symmetric games: A mean field theoretic approach" in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, 2019), pp. 12134-12144.
50. P. Mertikopoulos, W. H. Sandholm, Learning in games via reinforcement and regularization. *Math. Oper. Res.* **41**, 1297-1324 (2016).
51. T. Li, G. Peng, Q. Zhu, T. Başar, The confluence of networks, games, and learning a game-theoretic framework for multiagent decision making over networks. *IEEE Control Syst. Mag.* **42**, 35-67 (2022).
52. Y. Sato, J. P. Crutchfield, Coupled replicator equations for the dynamics of learning in multiagent systems. *Phys. Rev. E* **67**, 015206 (2003).
53. A. Kianercy, A. Galstyan, Dynamics of Boltzmann Q learning in two-player two-action games. *Phys. Rev. E* **85**, 041145 (2012).
54. N. Masuda, H. Ohtsuki, A theoretical analysis of temporal difference learning in the iterated Prisoner's dilemma game. *Bull. Math. Biol.* **71**, 1818-1850 (2009).
55. S. Dridi, L. Lehmann, On learning dynamics underlying the evolution of learning rules. *Theor. Popul. Biol.* **91**, 20-36 (2014).

56. T. Börgers, R. Sarin, Learning through reinforcement and replicator dynamics. *J. Econ. Theory* **77**, 1–14 (1997).
57. K. Tuyls, A. Nowé, Evolutionary game theory and multi-agent reinforcement learning. *Knowl. Eng. Rev.* **20**, 63–90 (2005).
58. W. Barfuss, Dynamical systems as a level of cognitive analysis of multi-agent learning. *Neural Comput. Appl.* **34**, 1653–1671 (2022).
59. M. Kaisers, D. Bloembergen, K. Tuyls, "A common gradient in multi-agent reinforcement learning" in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems—Vol. 3, AAMAS '12* (2012), pp. 1393–1394.
60. M. Pangallo, T. Heinrich, J. Doyne Farmer, Best reply structure and equilibrium convergence in generic games. *Sci. Adv.* **5**, eaat1328 (2019).
61. L. E. Blume, The statistical mechanics of strategic interaction. *Games Econ. Behav.* **5**, 387–424 (1993).
62. D. C. Krakauer, J. C. Flack, Better living through physics. *Nature* **467**, 661 (2010).
63. N. E. Leonard, S. A. Levin, Collective intelligence as a public good. *Collect. Intell.* **1**, 26339137221083293 (2022).
64. B. Walker, C. S. Holling, S. R. Carpenter, A. Kinzig, Resilience, adaptability and transformability in social-ecological systems. *Ecol. Soc.* **9**, 5 (2004).
65. J. F. Donges, W. Barfuss, From math to metaphors and back again: Social-ecological resilience from a multi-agent-environment perspective. *GAIA: Ecol. Persp. Sci. Soc.* **26**, 182–190 (2017).
66. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554–2558 (1982).
67. P. Stone, G. Kaminka, S. Kraus, J. Rosenschein, "Ad hoc autonomous agent teams: Collaboration without pre-coordination" in *Proceedings of the AAAI Conference on Artificial Intelligence* (2010), vol. 24, pp. 1504–1509.
68. F. P. Santos, J. M. Pacheco, A. Paiva, F. C. Santos, "Evolution of collective fairness in hybrid populations of humans and agents" in *Proceedings of the AAAI Conference on Artificial Intelligence* (2019), vol. 33, pp. 6146–6153.
69. W. Barfuss, J. F. Donges, V. V. Vasconcelos, J. Kurths, S. A. Levin, Caring for the future can turn tragedy into comedy for long-term collective action under risk of collapse. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12915–12922 (2020).
70. A. Lazaric, E. Munoz, F. de Cote, M. Restelli Dercole, "Bifurcation analysis of reinforcement learning agents in the Selten's horse game in adaptive agents and multi-agent systems III" in *Adaptation and Multi-Agent Learning*, K. Tuyls, A. Nowé, Z. Guessoum, D. Kudenko, Eds. (Springer, Berlin/Heidelberg, Germany, 2008), pp. 129–144.
71. M. Kaisers, K. Tuyls, "FAQ-learning in matrix games: Demonstrating convergence near Nash equilibria, and bifurcation of attractors in the battle of sexes" in *Proceedings of the 13th AAAI Conference on Interactive Decision Theory and Game Theory, AAAIWS'11-13* (AAAI Press, 2011), pp. 36–42.
72. S. Leonardos, G. Piliouras, Exploration-exploitation in multi-agent learning: Catastrophe theory meets game theory. *Proc. AAAI Conf. Artif. Intell.* **35**, 11263–11271 (2021).
73. M. Scheffer et al., Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).
74. D. H. Wolpert, M. Harré, E. Olbrich, N. Bertschinger, J. Jost, Hysteresis effects of changing the parameters of noncooperative games. *Phys. Rev. E* **85**, 036102 (2012).
75. G. Yang, D. Basanta, G. Piliouras, Bifurcation mechanism design—From optimal flat taxes to better cancer treatments. *Games* **9**, 21 (2018).
76. X. Chen, F. Fu, Imperfect vaccine and hysteresis. *Proc. R. Soc. B* **286**, 20182406 (2019).
77. S. Leonardos, G. Piliouras, Exploration-exploitation in multi-agent learning: Catastrophe theory meets game theory. *Artif. Intell.* **304**, 103653 (2022).
78. S. Wieczorek, C. Xie, P. Ashwin, Rate-induced tipping: Thresholds, edge states and connecting orbits. *Nonlinearity* **36**, 3238 (2023).
79. N. Creanza, O. Kolodny, M. W. Feldman, Cultural evolutionary theory: How culture evolves and why it matters. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7782–7789 (2017).
80. J. Newton, Evolutionary game theory: A renaissance. *Games* **9**, 31 (2018).
81. M. Botvinick, J. X. Wang, W. Dabney, K. J. Miller, Z. Kurth-Nelson, Deep reinforcement learning and its neuroscientific implications. *Neuron* **107**, 603–616 (2020).
82. A. G. Barto, S. Singh, N. Chentanez, "Intrinsically motivated learning of hierarchical collections of skills" in *Proceedings of the 3rd International Conference on Development and Learning* (Piscataway, NJ, 2004), pp. 112–119.
83. E. Hughes et al., "Inequity aversion improves cooperation in intertemporal social dilemmas" in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2018), vol. 31.
84. A. Modirshanechi, J. Brea, W. Gerstner, A taxonomy of surprise definitions. *J. Math. Psychol.* **110**, 102712 (2022).
85. O. Mihaescu, R. Neuneier, Risk-sensitive reinforcement learning. *Mach. Learn.* **49**, 267–290 (2002).
86. A. Klyubin, D. Polani, C. Nehaniv, "Empowerment: A universal agent-centric measure of control" in *2005 IEEE Congress on Evolutionary Computation* (IEEE, Edinburgh, UK, 2005), vol. 1, pp. 128–135.
87. K. J. Friston, J. Daunizeau, S. J. Kiebel, Reinforcement learning or active inference? *PLoS ONE* **4**, e64211 (2009).
88. M. A. Wilson, B. L. McNaughton, Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994).
89. A. Radulescu, Y. S. Shin, Y. Niv, Human representation learning. *Annu. Rev. Neurosci.* **44**, 253–273 (2021).
90. J. Schriftwieser et al., Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature* **588**, 604–609 (2020).
91. W. Barfuss, "Reinforcement learning dynamics in the infinite memory limit" in *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '20* (International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2020), pp. 1768–1770.
92. W. Barfuss, J. M. Meylahn, Intrinsic fluctuations of reinforcement learning promote cooperation. *Sci. Rep.* **13**, 1309 (2023).
93. W. Barfuss, R. P. Mann, Modeling the effects of environmental and perceptual uncertainty using deterministic reinforcement learning dynamics with partial observability. *Phys. Rev. E* **105**, 034409 (2022).
94. T. Galla, Intrinsic noise in game dynamical learning. *Phys. Rev. Lett.* **103**, 198702 (2009).
95. A. Traulsen, J. C. Claussen, C. Hauert, Coevolutionary dynamics: From finite to infinite populations. *Phys. Rev. Lett.* **95**, 238701 (2005).
96. M. A. Nowak, A. Sasaki, C. Taylor, D. Fudenberg, Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).
97. L. Zheng et al., "MAgent: A many-agent reinforcement learning platform for artificial collective intelligence" in *Proceedings of the AAAI Conference on Artificial Intelligence* (2018), vol. 32.
98. W. Wang et al., "From few to more: Large-scale dynamic multiagent curriculum learning" in *Proceedings of the AAAI Conference on Artificial Intelligence* (2020), vol. 34, pp. 7293–7300.
99. S. Hu, C. W. Leung, H. f. Leung, H. Soh, "The dynamics of Q-learning in population games: A physics-inspired continuity equation model" in *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS '22* (International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2022), pp. 615–623.
100. C. w. Leung, S. Hu, H. f. Leung, "Modelling the dynamics of multi-agent Q-learning: The stochastic effects of local interaction and incomplete information" in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence* (International Joint Conferences on Artificial Intelligence Organization, Vienna, Austria, 2022), pp. 384–390.
101. Z. Wang, C. Mu, S. Hu, C. Chu, X. Li, "Modelling the dynamics of regret minimization in large agent populations: A master equation approach" in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence* (International Joint Conferences on Artificial Intelligence Organization, Vienna, Austria, 2022), pp. 534–540.
102. C. Chu et al., "A formal model for multiagent Q-learning dynamics on regular graphs" in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence* (International Joint Conferences on Artificial Intelligence Organization, Vienna, Austria, 2022), pp. 194–200.
103. M. Laurière et al., Learning mean field games: A survey. *arXiv [Preprint]* (2022). <https://doi.org/10.48550/arXiv.2205.12944> (Accessed 12 June 2024).
104. Y. Yang et al., "Mean field multi-agent reinforcement learning" in *Proceedings of the 35th International Conference on Machine Learning*, J. Dy, A. Krause, Eds. (PMLR, 2018), pp. 5571–5580.
105. P. Ashcroft, P. M. Altrock, T. Galla, Fixation in finite populations evolving in fluctuating environments. *J. R. Soc. Interface* **11**, 20140663 (2014).
106. C. S. Gokhale, C. Hauert, Eco-evolutionary dynamics of social dilemmas. *Theor. Popul. Biol.* **111**, 28–42 (2016).
107. A. Szolnoki, X. Chen, Environmental feedback drives cooperation in spatial social dilemmas. *Europhys. Lett.* **120**, 58001 (2017).
108. A. R. Tilman, J. B. Plotkin, E. Alçay, Evolutionary games with environmental feedbacks. *Nat. Commun.* **11**, 915 (2020).
109. C. Hauert, C. Saade, A. McAvoy, Asymmetric evolutionary games with environmental feedback. *J. Theor. Biol.* **462**, 347–360 (2019).
110. J. Perolat et al., "A multi-agent reinforcement learning model of common-pool resource appropriation" in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17* (Curran Associates Inc., Red Hook, NY, 2017), pp. 3646–3655.
111. D. Hennes, M. Kaisers, K. Tuyls, "RESQ-learning in stochastic games" in *Adaptive and Learning Agents Workshop at AAMAS* (2010), p. 8.
112. M. P. Wellman, "Methods for empirical game-theoretic analysis" in *Proceedings of the 21st National Conference on Artificial Intelligence—Volume 2, AAAI'06* (AAAI Press, Boston, MA, 2006), pp. 1552–1555.
113. A. Vezhnevets, Y. Wu, M. Eckstein, R. Leblond, J. Z. Leibo, "Options as responses: Grounding behavioural hierarchies in multi-agent reinforcement learning" in *International Conference on Machine Learning* (PMLR, 2020), pp. 9733–9742.
114. W. Barfuss, "Supplementary information to collective cooperative intelligence". <https://github.com/wbarfuss/collective-cooperative-intelligence>. Zenodo. <https://doi.org/10.5281/zenodo.11611242>. Deposited 24 June 2024.