Cooperative control of environmental extremes by artificial intelligent agents

Martí Sánchez-Fibla, Clément Moulin-Frier, Ricard Solé

Cooperative control of environmental extremes by artificial intelligent agents

Martí Sánchez-Fibla[1,2], Clément Moulin-Frier[3] and Ricard Solé[4,5,6*]
[1] *Department of Information and Communications Technologies, Universitat Pompeu Fabra, 08018, Barcelona, Spain*
[2] *Artificial Intelligence Research Institute, IIIA (CSIC), Campus de la UAB, Bellaterra, Barcelona, 08193, Spain*
[3] *Inria - Flowers team Universite de Bordeaux ENSTA ParisTech*
[4] *Complex Systems Lab, Universitat Pompeu Fabra, Dr Aiguader 88, 08003 Barcelona, Spain*
[5] *Institució Catalana de Recerca i Estudis Avançats, Lluis Companys 23, 08010 Barcelona, Spain and*
[6] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA*
(Dated: December 27, 2024)

Humans have been able to tackle biosphere complexities by acting as ecosystem engineers, profoundly changing the flows of matter, energy and information. This includes major innovations that allowed to reduce and control the impact of extreme events. Modelling the evolution of such adaptive dynamics can be challenging given the potentially large number of individual and environmental variables involved. This paper shows how to address this problem by using fire as the source of extreme events. We implement a simulated environment where fire propagates on a spatial landscape, and a group of artificial agents learn how to harvest and exploit trees while avoiding the damaging effects of fire spreading. The agents need to solve a conflict to reach a group-level optimal state: while tree harvesting reduces the propagation of fires, it also reduces the availability of resources provided by trees. It is shown that the system displays two major evolutionary innovations that end up in an ecological engineering strategy that favours high biomass along with the suppression of large fires. The implications for potential A.I. management of complex ecosystems are discussed.

I. INTRODUCTION

The term "extreme event" is becoming a common description of a broad class of unanticipated natural events that can have disproportionate social, economic and ecological impacts. This term has been used in very different contexts, including mass extinctions [1], earthquakes and volcanic eruptions, and other natural hazards [2] as well as economic crashes [3]. In all these cases, the events occur over a very short time scale compared with that of the baseline dynamics. Because of how they can impact our lives, predicting these unlikely events has been a major source of research. The ability to predict an event boils down to whether the underlying system is governed by deterministic rather than stochastic dynamical patterns. If predictability is limited, an alternative path to deal with these rare events is to adapt to them using active strategies that reduce their impact or even suppress them. In this paper, we explore the latter.

Because of the accelerated pace of climate change, mega-fires, devastating floods and droughts jeopardise essential services and infrastructures, from agriculture to biodiversity. These events are expected to become more common in the coming decades [4]. Along with changes in energy use, novel agroforestry practices and conservation policies, intervention scenarios also need to be considered [5] that take into account the complex, multiscale nature of the problem in space and time [6, 7].

*Corresponding author: ricard.sole@upf.edu

The uncertainty associated with environmental fluctuations is far from new to humans. Our ecological success is due to a combination of features favouring developing a culturally evolved cooperative social environment [8]. In this way, humans became unique in interacting with the environment. Tool making and social intelligence paved the way for an unprecedented transformation of the biosphere, with humans becoming large-scale *ecosystem engineers* [9] i. e. a species having a major impact on the flows of energy and matter [10]. Agriculture for example, can be understood as a powerful way of reducing environmental uncertainty [11]. Similarly, the emergence of urban environments profoundly changed our relationship with nature and its uncertainties [12]. How do these major innovations occur? Moreover, for a given fluctuating environment, what role does learning play in finding efficient outcomes?

A central problem, in general, is finding emergent solutions to environmental challenges, such as resource scarcity and the impact of extreme events. A range of theoretical approaches, including agent-based models [13–15], computational ecologies [16–18], game theory [19–21] to statistical physics [22], have shown that robust adaptive behaviour emerges from conflicting constraints [23, 24]. Moreover, the study of future challenges associated to climate change has also benefited from models that include humans and the environment altogether [26, 70] and allow to understand the potential transitions between dynamical states [27]. While some of these models consider constant parameters, a great potential for finding adaptive solutions come from coevolution between agents and their environments [28, 29]. Can such evolution between agents and their environment allow

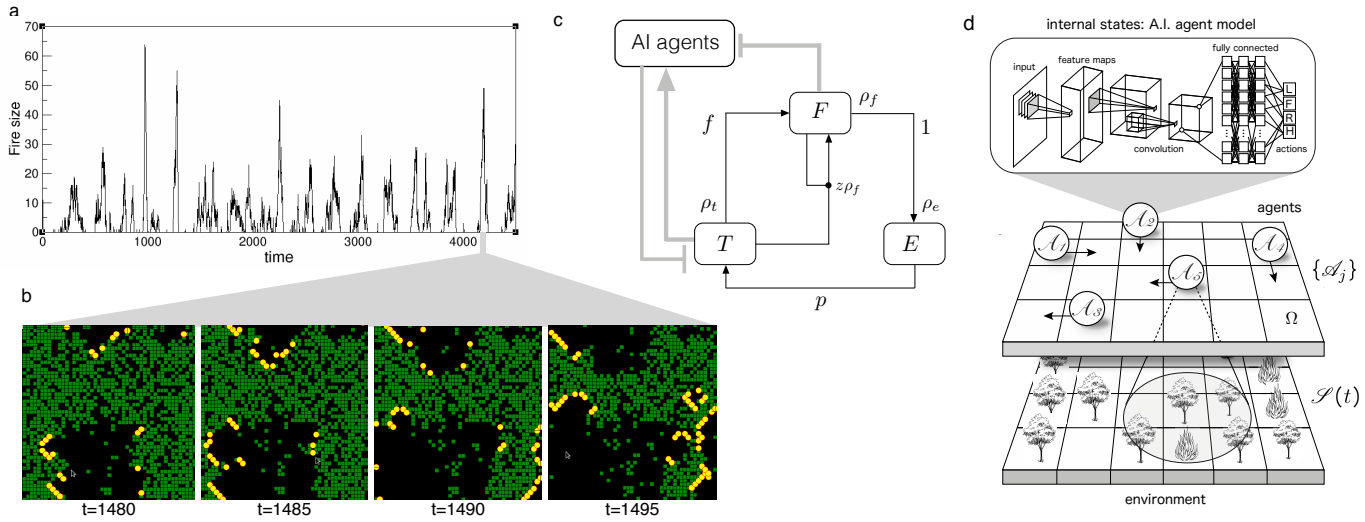

FIG. 1: Forest fire dynamics in time and space. In figure (a) a typical time series of the number of burned sites in a forest fire model (FFM) is displayed for a square lattice with $L = 50$ and parameters $p = 0.003, f = 0.00003$. The number of sites burning (the fire size) shows marked bursting dynamics. Four spatial snapshots are shown in (b) associated with a fire burst. Here, green, yellow and black correspond to trees, fires and ashes (empty sites) respectively. The basic set of rules is summarized in (c) using black arrows. In our model, we add a set of AI agents whose interactions with the environment are marked with grey arrows where positive and negative interactions are indicated as $\rightarrow$ and $\dashv$, respectively. They benefit from trees but get punished by fire spreading, and can modify tree density by harvesting trees. In (d) we summarize the levels of interaction between forest fire dynamics and its control by neural agents. The bottom layer defines the observed spatial pattern of states of the Forest Fire Model (FFM), which changes stochastically while can be affected by the action of agents (middle layer) that have a limited observation range (indicated as a circle in the bottom layer) and can take decisions about their movement and harvesting trees locally. Each agent (upper layer) makes decisions (implements an action policy, mapping observed states to actions) by means of a convolutional neural network trained with Reinforcement Learning (RL). The RL process eventually defines the behavioural pattern displayed by the agent, which translates into a set of potential actions $(LeftTurn, RightTurn, ForwardMove, Harvest)$ in response to the local environment.

finding strategies to deal with environmental extremes?

In this paper, we approach the previous questions by considering a scenario where a set of artificial learning agents interact with a forest where large fires can occur. The agents can learn to exploit a finite set of resources (gathered from trees) while dealing with the destructive potential of forest fires (that require trees to propagate). The dilemma here is quite obvious: a high fraction of tree cover gives larger opportunities for harvesting but also allows for large fires to occur. What kind of strategy can balance this conflict? The case study chosen here allows us to clearly define the constraints imposed by the environment and the repertoire of tasks to be performed by the agents. The choice of fire is grounded in its ecological impact on a wide range of ecosystems all over the planet and its role in human history [30, 31]. Fire burns ecosystems acting as a major evolutionary force. At the same time, its use by humans is connected to a major innovation used by our species as a way of engineering the wild. Here we approach the problem by using a hybrid model that combines forest fire spreading with a set of artificial learning agents that adapt to environmental conditions and exploit resources while finding ways to control fire. The outcome of the evolution of these agents is the emergent cooperative management of flammable ecosystems that provides both a higher tree yield together with a marked fire reduction.

Agents will evolve as they interact and modify fire behaviour using reinforcement learning (RL) [32]. RL is one of the key directions of Machine Learning and is playing a major role in ongoing developments within Artificial Intelligence (AI, see [33]), spanning multiple domains, from game theory to advanced robotic tasks. In a nutshell, in RL, optimal strategies emerge within the lifetime of an agent (both learner and decision-maker) who interacts with an environment, giving rise to (cumulative) rewards that the agent learns to maximise. When considered in a multi-agent context, where multiple RL agents interact in a shared environment (Multi-Agent Reinforcement Learning, MARL, [34]), RL provides a powerful approach to exploring social dilemmas [35] or the learning and maintenance of social norms within societies [36]. As shown below, cooperative strategies emerge as agents deal with fire, evolving different forms of exploiting resources while protecting themselves from damage, first using a simple herdingstrategy and later on developing a more sophisticated one where decision-making depends on a distributed management of the local environment.

Modelling the evolution of ecological control by a population of agents requires two main components. The first is a model of extreme environmental fluctuations provided by fire spread. We implement it using a minimalist Forest Fire Model (FFM, [40, 41]) framework based on cellular automata (CA), which have been successfully used in many areas [37–39]. The second is the formalization of a set of agents that learn how to respond to these fluctuations while gathering resources from trees. We implement it as a population of independent RL agents navigating in a shared FFM environment and learning from experience how to control behavior, without having access to the observations, actions and rewards of the others (decentralized MARL, [34–36]).

Spatial dynamics of fire. Fire spread is described by means of the *Forest Fire Model* [40, 41] where a toy description of fire spreading is applied to the states of each site on a two-dimensional, $L \times L$ lattice Ω. For convenience we use periodic boundary conditions (i. e. dynamics takes place on a torus). Each site $\mathbf{r} \in \Omega \in \mathbb{Z}^2$ can be in three possible states, namely $S(\mathbf{r}) \in \Sigma = \{E, T, F\}$, where E denotes an empty cell, T a tree cell and F a fire cell. The state of the system, $\mathcal{S}$ will be updated by means of three probabilistic events, namely: (1) spontaneous burning of a tree, i. e. a transition

$$T \xrightarrow{f} F$$

at a rate f, leading to a burning site (fire cell); (2) growth of new trees from empty sites, i. e. with a probability p we have

$$E \xrightarrow{p} T.$$

(3) The last rule allows fire propagation: if a given tree has a neighbour that is a fire, it burns too. This means a (deterministic) transition

$$T \xrightarrow{1} F.$$

Hereafter, the set of neighbours $\Gamma(\mathbf{r})$ is defined by a *von Neumann neighborhood*, i. e. the four nearest ones. For our two-dimensional lattice, we have $\mathbf{r} = (i, j)$ and $\Gamma(\mathbf{r}) = \{(i \pm 1, j), (i, j \pm 1)\}$.

It can be shown (see SM section I, where SM refers to the Supplementary Material) that the previous discrete rules allow to define a mathematical model of forest fire dynamics that converges to a stable attractor (fixed point) where fires and trees coexist. However, for $p, f \ll 1$ and $f \ll p$ the actual discrete, spatially-explicit dynamics is highly fluctuating, exhibiting a broad spectrum of fluctuations (a self-organized critical (SOC) state) that includes extreme events [42, 43]. In figure 1a-b we show an example of the fire spreading dynamics on a $L = 50$ lattice with four snapshots associated to a major fire event. The origins of such extreme events are to be found in the separation of time scales associated to the SOC dynamics [41, 44]. Despite of its simplicity, the FFM and variations of it has been successfully applied to model the statistical patterns of the actual fires [45–49]. Other similar models that exhibit SOC have been used to study other systems displaying extreme phenomena, including earthquakes, rainfall patterns, fractal river networks or financial markets [42, 50, 51].

The basic rules and transitions are summarized in figure 1*c*, along with the schematic interaction with agents, which benefit from the presence of trees but need to avoid the damage caused by fires. Because of the fluctuating nature of fires (which can exhibit large peaks of destruction, see figure 1*b*), resource availability and potential damage by fire can be rather unpredictable. It can be shown (see SM sections I,2) that the previous discrete rules allow to define a mathematical model of forest fire dynamics that converges to a stable attractor (fixed point) where fires and trees coexist.

Agent-environment interaction. We consider a population of N learning agents $\{\mathcal{A}_j\}$ with $j = 1, ..., N$, interacting with the environment as defined by $\mathcal{S}(t)$. Initially ($t = 0$), each agent is placed randomly on Ω and given a random orientation (either North, East, South or West). Each cell of the grid-world environment can contain one agent (fig. 1d) which can influence the state of $\mathcal{S}(t)$ by executing some actions, as defined below. In order to describe the profit tied to the exploitation of trees, we introduce a resource associated to a tree that can be a source of reward for the learning agent. Specifically, a tree can carry a resource (say a fruit) that can be consumed by agents. Let us indicate as T^* these fruit-carrying tree. Once consumed, it can be restored after some time, given by a recovery rate. This extra state does not modify the fire dynamics, since it does not affect the FFM rules.

At each time step, each agent can randomly choose to move forward (a_F) to a neighbouring site if not occupied by another agent, rotate to the left (a_L), rotate to the right (a_R), or harvest the site in front of it (a_H). If, after moving, the new site is a resource it will be consumed by the agent, i.e it will become a tree cell (which will be able to regenerate a resource with probability P_r). The harvesting a_H action will only have an effect if the cell in front of the agent is occupied by a tree with no available resources. In that case, the tree will be removed (i. e. replaced by an empty cell) making the transition:

$$T \in \Gamma(\mathbf{r}) \xrightarrow{\mathcal{A}} E$$

An agent consuming a resource will receive a positive reward R_r, while an agent residing on a fire cell will receive a negative reward R_f (with R_r much smaller than R_f). We can summarize these extra transitions as follows:

$$T^* \xrightarrow{\mathcal{A}} T \xrightarrow{P_r} T^*$$

where the first transition indicates that the presence of an agent implies the loss of the resource. The agent-environment interaction dynamics is formalized as a Partially Observable Markov Decision Process (POMDP) in

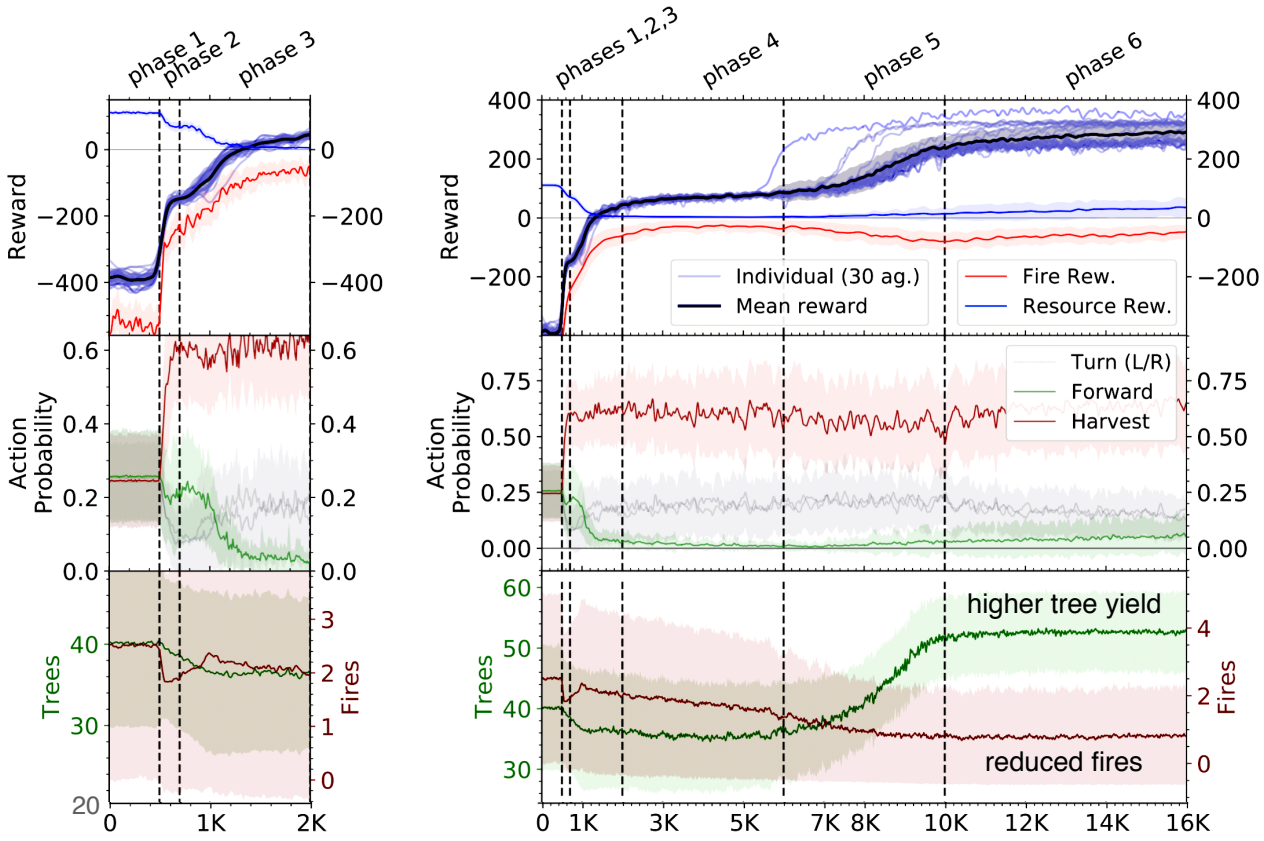

FIG. 2: Evolving cooperation by artificial learning agents. The time series of all the relevant measures of a population of RL agents that evolve ecological engineering strategies is shown over $16K$ episodes. This includes the reward for all agents, mean action probability usage of all agents, number of trees (mean and standard deviation), number of fires (mean and standard deviation). The left column is a zoomed view of the first $2K$ episodes, with 500 time steps each. Parameters are: $p_t = 0.08, p_f = 0.005$ for the FFM and $L = 10, N = 30$ with a 7×7 observation window. On the right the full time series is displayed. Notice the marked increase of tree yield and the suppression of fires, which also involves a drastic reduction of fluctuations.

SM section IV and is illustrated in figure S2 (also in the SM).

Agent learning. The objective of each agent is to learn an action policy maximizing its own reward, i.e. to maximize the number of collected resources while avoiding to be burned by fire. Learning is structured in a sequence of episodes, with a fixed duration of T time steps each. At the start of each episode, a new map is randomly initialized with tree and empty cells according to a probability distribution p_{init} and N agents are randomly positioned on the map (random positions and orientations).

Each agent learns its own action policy, mapping its partial observation of the environment at the current time step to a probability distribution over its actions. At each time step, each agent only observes its own local neighborhood (see SM section III for details). The learning objective is to maximize the cumulative reward obtained over an entire episode:

$$G_t = \sum_{t=0}^{T-1} \gamma^t R_t \tag{1}$$

where R_t is the reward obtained by the agent at time step t and $\gamma < 1$ is a discount factor.

The action policy described above is generated by a 2D Convolutional Neural Network in which every agent trains independently in order to maximize its cumulative reward G_t (see SM section V for all network details and table I for all parameters used). The network weights are regularly updated from the agent's experience, i. e. from the tuples (state, action, next state and reward) collected at each time step. The agents act randomly at the start given the random initialization of weights. During training they are able to maintain a certain level of exploration by favoring (with a small contribution) the training loss towards the equiprobable distribution of actions.

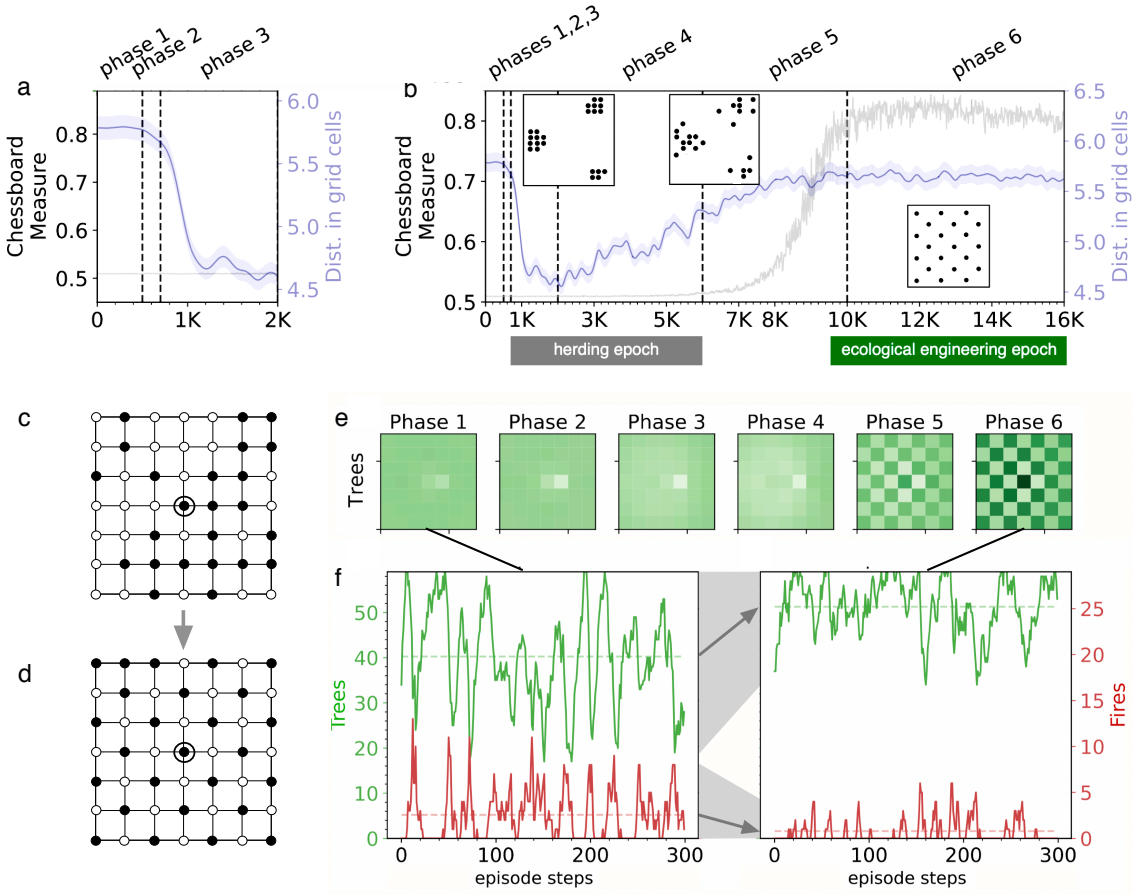

FIG. 3: Characterization of the observed cooperative phase transitions. In (a-b), the time series for the herding (blue) and ecological-engineering (grey) measures are displayed. The herding measure characterizes the agents' tendency to form dense herds. It is illustrated in the three insets in (b), sketching typical spatial arrangements of agents: a high herding measure indicates that agents form dense herds (left inset);. In contrast, lower herding measures indicate that they are more uniformly spread in the environment (middle and right inset). The ecological-engineering measure characterizes the agents' ability to create a structured pattern of trees limiting fire propagation. Patterns of trees corresponding to a low and high ecological-engineering measure are illustrated in (c) and (d), respectively. The open circle in the middle indicates an agent, with black and white circles indicating the presence or absence of trees, respectively. Intuitively, the high measure resulting from (d) corresponds to a perfect chessboard pattern preventing fire propagation (which only propagates in the horizontal and vertical dimensions) while maximizing the number of trees. The progressive formation of this structured pattern of trees is displayed in (e), showing the average density of trees in all agent's neighbourhoods during the six emerging phases (in the SM figure S4 we show all agents' observations grids). In (f), we observe the FFM dynamics in the first and last episodes of the simulation. This demonstrates that the agent population managed to increase (resp. decrease) the average number of trees (resp. fire) as indicated by the horizontal dotted lines while reducing the fluctuation range of trees and fires.

II. RESULTS

Under the environmental description provided by the system's state $\mathcal{S}$, the agents benefit from tree-rich neighbourhoods since trees provide the resources and reward. But trees also propagate fire and agents are highly punished when being burned (through a negative reward), thus creating an incentive to harvest the trees around them. The agents must, therefore, learn a strategy that maximizes the number of trees around them while minimizing fire propagation. How will the agents solve this conflict? As shown below, several rapid changes in agent behaviour occur, involving two marked cooperative transitions. The model reveals several phases of evolution where agents learn to protect themselves from fire first by clustering in groups (the herding epoch) while, in the long run, ecological engineering is developed as a spatially distributed decision-making pattern (the ecological engineering epoch).

The results reported in this section correspond to a simulation of 16k episodes of $T = 500$ steps each with $N = 30$ agents in a 10×10 grid and the FFM model having fixed probabilities of tree regrowth $p_{tree} = 0.08$ and fire appearance $p_{fire} = 0.008$. In SM section VII we did

a full hyper-parameter analysis of these two environmental parameters. In SM figure S3 we show screenshots of bigger simulations.

Figure 2 (right) shows the six phases which spontaneously emerge from the interaction between fire spreading and learning dynamics. Agent's learning only starts at the end of phase 1 (after 500 episodes), before which random actions occur and low rewards are observed (as expected). In this phase, the number of trees and fires are consistent with the FFM predictions (see SM section I). Training starts at phase 2, where a rapid increase in harvesting (and a decrease in the rotation actions) is observable, while maintaining the forward action approximately in order to collect resources. The mean reward raises consequently, yet is still highly negative, indicating that foraging and harvesting are not performed efficiently (fig.2, left).

Then, from phase 3, cooperative transitions start to emerge. In order to characterize them, we propose quantitative measures that we explain here intuitively and define formally in the SM. First, we define a herding ~~m~~easure characterizing the tendency of the agent population to form dense herds (computed as the average density of other agents in each agent's neighborhood). It corresponds to the blue curve in figure 3a-b and is illustrated in the insets of figure 3b. Second, we define a measure of ecological engineering characterizing the agents' ability to create a structured pattern of trees limiting fire propagation while maximizing the number of trees. It corresponds to the grey curve in figure 3a-b and is illustrated in the figure 3c-d.

Herding phase. Using the herding measure, we can detect a first cooperative change, which we label the *herding epoch*. It is characterized by high values of this measure, constantly increasing and reaching a maximum from episodes 1000 to 2000 (phase 3) as seen in figure 3b. As observed in figure 2 a sudden drop in the forward action takes place along with an increase in the rotation actions, while harvesting occurs at a high rate (around 60% of the time). These measures indicate that the agents learn to rapidly form a packed group at the beginning of the episode, then stay in place while harvesting trees around them, creating a safe area where they are efficiently protected from fire. As a consequence, the proportion of fire cells decreases (but is still relatively high due to fire propagation outside of the group area). We observe that the negative rewards received when agents are burned by fire and the positive rewards from collected resources both approach 0, indicating that the agents are well protected from fire but are not able to collect resources within the packed group. While sub-optimal, this herding strategy still makes sense at this stage since the penalty for getting burned R_f is much stronger than the positive reward R_r of consuming a resource.

Expansion phase. In the next phase in the evolution of agent behaviour (phase 4, episodes 2000 to 6000), the agents improve upon the herding~~s~~trategy discovered in the previous phase. While reward grows (figure 2) they maintain the group coherence while allowing more space between agents within the group, as indicated by the decrease in herding measure (figure 3): clusters start to expand in space, occupying larger areas. The proportion of fire cells in the environment continues to decrease, since the area of the grid covered by the group increases and the spaces newly introduced between the agents are too small for fire to propagate. These observations indicate that a new collective strategy is building up. The aftermath of the next phase (episodes 7000 to 10000, phase 5) is marked by an increase in reward while agents become more isolated, along with a reduction of fires and an increase in tree cover. This phase defines the transient towards a new phase characterized by the dominance over fluctuations and the development of fire suppression. How is this achieved? This transition is the result of a spatially-extended control of fire spread resulting from a new set of decisions based on a more accurate control of the agents over their environment. This can be quantified by means of the ecological engineering measure. The motivation of this measure, which intuitively indicates the "chessboardness" of the pattern of trees (figure 3d), is rooted in the emergence of a behavioral pattern where active harvesting of most close four neighboring trees occurs, while those in the four diagonals (where fire cannot propagate) tend to be free from harvesting.

In this fifth phase, we observe a sudden increase in the ecological engineering measure, indicating that the agents learn to harvest trees in a much more structured way. The resulting chessboard pattern is predicted as an optimal structure that can prevent fire propagation while allowing trees (and therefore resources) to appear on half of the grid cells. The effectiveness of this pattern is confirmed by a substantial increase in the proportion of trees in the grid, even though the harvest action continues to be executed at the same rate as in the previous phase. We also observe that the proportion of fire cells continues to decrease, confirming the effectiveness of the chessboard pattern in reducing fire propagation, thus increasing the mean reward.

Ecological engineering phase. The last phase in the evolution of our RL agents (phase 6, episodes 10k to 16k) involves control over fire spread. We label it the *ecological~~ecosystem~~ engineering epoch* (marked with a colour bar in figure 3b). The ecological engineering measure remains constant at its maximum value in this sixth phase. The agents benefit from this well-engineered ecosystem, as shown by the convergence of the mean reward towards its maximum. However, we observe a significant variance in the reward, with a few agents obtaining much more reward than others. However, the higher reward received by these agents does not seem to negatively impact the reward of others, suggesting that they are more "risk-takers" than "free-riders", i. e. agents that take the risk of moving across the grid to collect more rewards without dramatically impacting the structure of the global chessboard pattern.

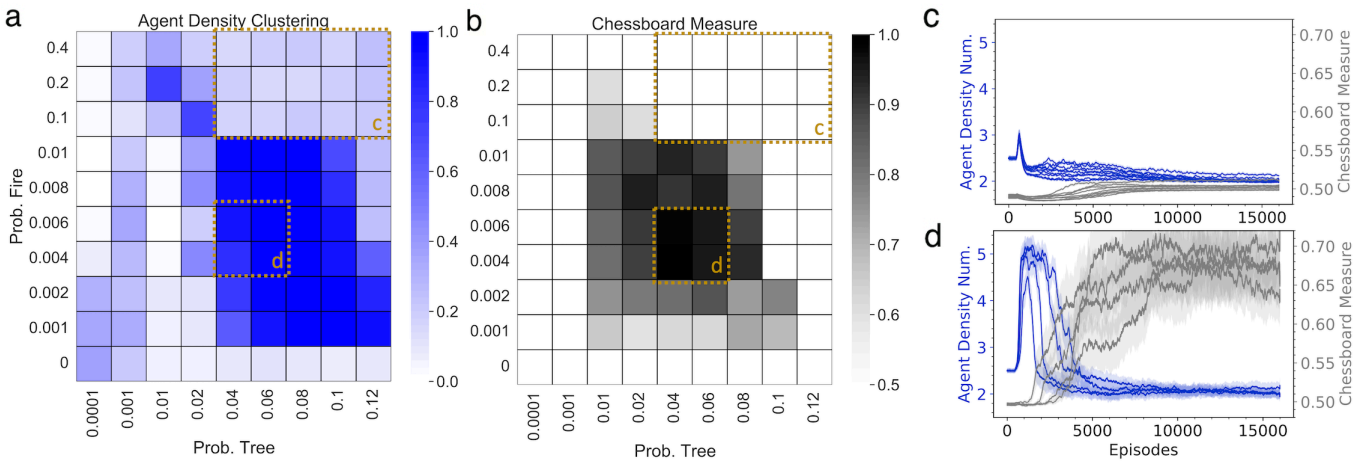

FIG. 4: The cooperation hyper-parameter space. The cooperation measures, agent herding as density clustering on panel (a) and ecological engineering of a chessboard pattern on panel (b) are plotted as a function of tree regrowth (x axis of the grids) and spontaneous fire appearance (y axis), which are the main hyper-parameter of the model, fixed throughout each simulation. Both grids in panels (a) and (b) provide an overview of 90 simulations of 16K episodes each. In (a), each cell shows the normalized mean agent density herding measure throughout the first 8K episodes of a simulation. In grid (b), the mean of the ecological engineering measure for the second half of the simulation is shown. Panels (c) and (d) show the episode dynamics of four simulations indicated in the grids by the yellow super-imposed dashed-dotted squares. Agent herding is maximum over a range of cells with a value of 0.71 (see figure S6). Ecological niche engineering only occurs in the central region of the space and collapses for p_{tree} less than 0.01 and greater than 0.08 and p_{fire} less than 0.001 and greater than 0.1. The simulations in panels (c,d) where chosen to put forward the facts that (c) complex niche engineering can occur without agent herding and (d) there is a region where both are maximum.

In figure 4 we take a deeper look into the emergence of the two identified cooperative strategies: agent grouping and chessboard arrangement (through the herding and ecological engineering measures, respectively). We study the robustness of cooperation for different environmental conditions. Cooperation is not widespread throughout all the parameter space of the FFM automata (tree regrowth and spontaneous fire appearance) as seen in the grids (a,b) of the figure were the maximum herding and ecological engineering measures are shown respectively. Cooperation collapses in all extremes: when there is no fire, when there is too much fire, when there are too few trees (too few resources) leading to competition and individualistic outcomes and when trees regrow rapidly facilitating fire propagation and fostering again individualistic strategies. Panels (c,d) of figure 4 further demonstrate that showing nonexistent cooperation throughout all the episodes of a large region of the parameter space compared to a high level of cooperation at the central region of the space (d) both in terms of clustering and chessboard formation (the simulation of figure 3 being a particular example of it). See the SM for a full hyper-parameter analysis on cooperation, figure S6.

Although cooperation is not widespread and collapses in the extremes of the parameter space, agents manage to control their environment in all situations by reducing fire propagation and increasing tree population (see SM figures S5,S6,S7).

III. DISCUSSION

In this paper, we have shown how multiagent learning dynamics is an important step towards understanding how can a collective of cognitive agents achieve cooperative control over extreme events. This is nowadays a timely issue, as global warming is rapidly disrupting the long temperature stability that allowed human civilization to thrive during the Holocene [52].

Human societies have been particularly successful in this respect by developing cooperation strategies [53]. Cooperation is thus a major force of nature control. But how such a control is achieved when conflicting constraints arise? Previous work has explored the dynamics of cooperation under game-theory approximations [19, 54], including those experiencing noisy conditions [55]. In other studies, the use of neural agents allowed to explore the emergence of simple cooperative strategies [56] and the interplay between cooperation and social intelligence [57]. However, our study is, to the best of our knowledge, the first work that considers the rise of cooperation among cognitive (RL) agents dealing with extreme events. The agents must develop novel, cooperative strategies when dealing with extreme events and conflicting constraints as those addressed here. Our results are in line with recent proposals on the role of ecosystem dynamics and niche construction in both biological and cultural evolution [58, 59] as well as in AI [60–62].

We have shown that the outcome of these conflicts

is several consecutive transitions that provide increasing opportunities to the agents, including two major events that reflect the partial and eventually global control of the entire ecosystem. This example illustrates the potential for A.I. systems to help explore novel ways to deal with the high-dimensional nature of complex environments. By suppressing fires, agents have effectively taken control of uncertainties while also obtaining stable resources with high yield. This illustrates the emergence of ecosystem engineering on a global scale. Moreover, our work shows how RL models can help to explore other human-ecological transitions under a synthetic approximation [63].

Several current extensions of the model can be considered for future work. In the current version, we have shown that several hyperparameters of the model have a significant influence on the collective learning dynamics. This is, for instance, the case of the spontaneous growth and burning probabilities of trees (p_{tree} and p_{fire}, respectively, as analysed in fig. 4). The effect of more environmental and learning hyperparameters of the model could be analysed in future work, all of them being provided in table I of the SM. We could also consider extending the agent's adaptive mechanisms. In the current version, the agent's adaptation is only driven by reward maximization through reinforcement learning (RL), i.e. we consider a developmental adaptation timescale. Bi-level optimisation algorithms, also called meta-learning, are increasingly used in the machine learning community. A particularly interesting approach is Meta Reinforcement Learning (Meta-RL) [64, 65], where an inner adaptive loop based on RL is itself meta-optimized by an outer loop operating at a larger timescale. In this sense, Meta-RL is sometimes considered as a model of how evolution shapes developmental learning in biological organisms, as a solution to adapt to a wide range of environmental conditions [62, 66]. An interesting direction for future work is, therefore, to apply Meta-RL to agent's populations that are exposed to various degrees of environmental variability, e.g. by randomizing environmental hyperparameters at each episode (e.g. p_{tree}, p_{fire}, or the presence of wind) and study if these conditions could result in the emergence of more generalist collective strategies. Finally, another perspective is to propose additional measures to evaluate the collective learning and eco-engineering dynamics in our model, e.g. based on information-theoretic measures such as environment entropy or agent empowerment [67].

Finally, although ours is a simple model, we believe that it illustrates the potential that A.I. systems not only to modelling and tackling the Earth system [68, 70, 71] or predict extreme events [69] but also to find novel solutions to the conservation, engineering and restoration of other ecosystems facing tipping points [72–75].

Within the specific context of forest fires, our toy model does not capture the true complexity of real wildfires (beyond the universal size distributions). More realistic cellular automata models have been developed to incorporate key variables such as weather conditions and topography [78, 79], wind conditions and fire suppression tactics [80] and even a Machine Learning determination of more detailed non-linear transformation rules for fire burning probabilities [81]. While all these models keep the cellular automaton description on a lattice, the repertoire of dynamical rules and the landscape heterogeneity define a set of realistic traits that should be tested to see if the collective solutions found by our RL agents are similar or are instead replaced by other strategies. Although still under development, the complex, spatially distributed, multiscale nature of ecosystems might require the help of A.I. systems capable of dealing with their emergent dynamics.

Acknowledgments

The authors thank the Complex Systems Lab members for fruitful discussions. Special thanks to Hari Seldon for his inspiring ideas. RS was supported by the Spanish Ministry of Economy and Competitiveness grant FIS2016-77447-R MINECO/AEI/FEDER, and the Santa Fe Institute. CMF was supported by the Inria Exploratory action ORIGINS (`https://www.inria.fr/en/origins`) as well as the French National Research Agency (`https://anr.fr/`, project ECOCURL, Grant ANR-20-CE23-0006). MSF was supported by the Spanish Ministry of Economy and Competitiveness grant DPI2016-80116-P MINECO/AEI/FEDER.

[1] Benton, M.J., 1995. Diversification and extinction in the history of life. Science, 268(5207), pp.52-58.

[2] Sachs, M.K., Yoder, M.R., Turcotte, D.L., Rundle, J.B. and Malamud, B.D., 2012. Black swans, power laws, and dragon-kings: Earthquakes, volcanic eruptions, landslides, wildfires, floods, and SOC models. The European Physical Journal Special Topics, 205, pp.167-182.

[3] Sornette, D., 2003. Critical market crashes. Physics reports, 378(1), pp.1-98.

[4] Peñuelas, J., Sardans, J., Estiarte, M. et al., 2013. Evidence of current impact of climate change on life: a walk from genes to the biosphere. Global change biology 19, 2303-2338.

[5] Solé, R. and Levin, S., 2022. Ecological complexity and the biosphere: the next 30 years. Phil. Trans. Royal Soc. B 377, 20210376.

[6] Levin, S.A., 2000. Multiple scales and the maintenance of biodiversity. Ecosystems 3, 498-506.

[7] Solé, R. and Bascompte, J. 2007. *Self-organization in complex ecosystems*. Princeton U. Press. Princeton, USA.

[8] Boyd, R. and Richerson, P.J., 2009. Culture and the evolution of human cooperation. Phil. Trans. R. Soc. B 364,

3281-3288.
[9] Vitousek PM et al (1997) Human domination of Earth ecosystems. Science 27: 494-499.
[10] Jones CG, Lawton JCG and Shachak M. 1994. Organisms as ecosystem engineers. Oikos 69: 373-386.
[11] Gowdy, J. and Krall, L., 2014. Agriculture as a major evolutionary transition to human ultrasociality. Journal of Bioeconomics, 16, 179-202.
[12] Maisels, C.K., 2003. *The emergence of civilization: From hunting and gathering to agriculture, cities, and the state of the near east*. Routledge.
[13] Epstein, J.M. and Axtell, R., 1996. *Growing artificial societies: social science from the bottom up*. MIT Press.
[14] Miller, J.H. and Page, S.E., 2009. *Complex adaptive systems: an introduction to computational models of social life*. Princeton University. Press.
[15] De Marchi, S. and Page, S.E., 2014. Agent-based models. Annual Review of political science, 17, 1-20.
[16] Huberman, B. A. (Ed.) 1988. *The ecology of computation* . Studies in Computer Science and Artificial Intelligence 2, North-Holland, Amsterdam, New York.
[17] Kephart, J.O., Hogg, T. and Huberman, B.A., 1989. Dynamics of computational ecosystems. Physical Review A, 40(1), p.404.
[18] Huberman, B.A., 1990. The performance of cooperative processes. Physica D: Nonlinear Phenomena, 42(1-3), 38-47.
[19] Axelrod, R. 2006. *The evolution of cooperation*. Basic Books, New York.
[20] Lindgren, K., 1991. Evolutionary phenomena in simple dynamics. Artificial life II, 10, 295-312.
[21] Nowak, M.A., 2006. Five rules for the evolution of cooperation. science, 314(5805), 1560-1563.
[22] Perc, M., Jordan, J.J., Rand, D.G., Wang, Z., Boccaletti, S. and Szolnoki, A., 2017. Statistical physics of human cooperation. Phys. Rep. 687, 1-51.
[23] Turchin, P. 2016. *Ultra Society*. Beresta Books.
[24] Lansing, J.S., 2003. Complex adaptive systems. Annual review of anthropology, 32(1), 183-204.
[25] Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K. et al. 2022. Tackling climate change with machine learning. ACM Computing Surveys 55(2), 1-96.
[26] Farahbakhsh, I., Bauch, C.T. and Anand, M., 2022. Modelling coupled humanâenvironment complexity for the future of the biosphere: strengths, gaps and promising directions. Phil. Trans. R. Soc. B 377(1857), 20210382.
[27] Bauch, C.T., Sigdel, R., Pharaon, J. and Anand, M., 2016. Early warning signals of regime shifts in coupled humanâenvironment systems. Proc. Natl. Acad. Sci. USA 113, 14560-14567.
[28] Miller, J.H., 1996. The coevolution of automata in the repeated prisoner's dilemma. Journal of Economic Behavior and Organization, 29(1), 87-112.
[29] Miller, J. H. 2022. *Ex-machina: coevolving machines and the origins of the social universe*. SFI Press. Santa Fe, NM.
[30] Goudsblom, J., 1992. The civilizing process and the domestication of fire. Journal of World History 3, 1-12.
[31] Bond, W.J. and Keeley, J.E., 2005. Fire as a global âherbivoreâ: the ecology and evolution of flammable ecosystems. Trends Ecol. Evol. 20, 387-394.
[32] R. Sutton, A. Barto, *Reinforcement learning: An introduction*, Cambridge, MIT press, 1998.
[33] Russell, S.J. and Norvig, P., 2016. *Artificial intelligence: a modern approach*. Prentice and Hall, New Jersey.
[34] Littman, M. Markov Games as a Framework for Multi-Agent Reinforcement Learning. *Machine Learning Proceedings 1994*. 157-163 (1994,1)
[35] Perolat, J., Leibo, J.Z., Zambaldi, V., Beattie, C., Tuyls, K. and Graepel, T., 2017. A multi-agent reinforcement learning model of common-pool resource appropriation. Advances in neural information processing systems, 30.
[36] KÃ¶ster, R., Hadfield-Menell, D., Everett, R., Weidinger, L., Hadfield, G.K. and Leibo, J.Z., 2022. Spurious normativity enhances learning of compliance and enforcement behavior in artificial agents. Proc. Natl. Acad. Sci. USA 119, e2106028118.
[37] Chopard, B. and Droz, M., 2000. *Cellular Automata Modelling of Physical Systems*.Cambridge University Press.
[38] Ilachinski, A., 2001. *Cellular automata: a discrete universe*. World Scientific Publishing Company.
[39] Batty, M., 2007. *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*. The MIT press.
[40] Bak, P., Chen, K. and Tang, C., 1990. A forest-fire model and some thoughts on turbulence. Physics letters A, 147(5-6), 297-300.
[41] Drössel, B. and Schwabl, F., 1992. Self-organized critical forest-fire model. Physical review letters 69, 1629.
[42] Bak P (1996) *How nature works. The science of self-organised criticality*. Copernicus, New York
[43] Turcotte, D.L., 1999. Self-organized criticality. Reports on progress in physics 62, 1377.
[44] Dickman, R., MuÃ±oz, M.A., Vespignani, A. and Zapperi, S., 2000. Paths to self-organized criticality. Brazilian Journal of Physics 30, 27-41.
[45] Malamud, B.D., Morein, G. and Turcotte, D.L., 1998. Forest fires: an example of self-organized critical behavior. Science 281, 1840-1842.
[46] Pueyo, S., 2007. Self-organised criticality and the response of wildland fires to climate change. Climatic Change 82, 131-161.
[47] Pueyo, S., De Alencastro GraÃ§a, P.M.L., Barbosa, R.I. eta l., 2010. Testing for criticality in ecosystem dynamics: the case of Amazonian rainforest and savanna fire. Ecology letters 13, 793-802.
[48] Van Nes, E.H., Staal, A., Hantson, S. et al., 2018. Fire forbids fifty-fifty forest. PloS one 13, e0191027.
[49] Hantson, S., Pueyo, S. and Chuvieco, E., 2015. Global fire size distribution is driven by human impact and climate. Global Ecology and Biogeography, 24, 77-86.
[50] Jensen, H.J., 1998. *Self-organized criticality: emergent complex behavior in physical and biological systems*. Cambridge U. Press. Cambridge UK.
[51] Rodriguez-Iturbe, I. and Rinaldo, A., 2001. *Fractal river basins: chance and self-organization*. Cambridge University Press. Cambridge UK.
[52] Ellis, E.C., 2015. Ecology in an anthropogenic biosphere. Ecological Monographs 85(3), 287-331.
[53] Carballo, D.M. ed., 2012. *Cooperation and collective action: archaeological perspectives*. University Press of Colorado.
[54] Roca, C.P., Cuesta, J.A. and Sanchez, A., 2009. Evolutionary game theory: Temporal and spatial effects beyond replicator dynamics. Physics of life reviews 6, 208-249.
[55] Helbing, D. and Yu, W., 2009. The outbreak of coopera-

tion among success-driven individuals under noisy conditions. Proceedings of the National Academy of Sciences 106, 3680-3685.

[56] Burtsev, M. and Turchin, P., 2006. Evolution of cooperative strategies from first principles. Nature 440, 1041-1044.

[57] McNally, L., Brown, S.P. and Jackson, A.L., 2012. Cooperation and the evolution of intelligence. Proceedings of the Royal Society B 279, 3027-3034.

[58] Odling-Smee, F., Laland, K. & Feldman, M. Niche construction: the neglected process in evolution. Princeton University Press.

[59] Laland, K., Uller, T. et al. The extended evolutionary synthesis: its structure, assumptions and predictions. *Proceedings Of The Royal Society B: Biological Sciences*. **282**, 20151019 (2015)

[60] Clune, J. AI-GAs: AI-generating algorithms, an alternate paradigm for producing general artificial intelligence. *ArXiv:1905.10985 [cs]*. (2020), http://arxiv.org/abs/1905.10985, arXiv: 1905.10985

[61] Leibo, J., Hughes, E., Lanctot, M. & Graepel, T. Autocurricula and the emergence of innovation from social interaction: A manifesto for multi-agent intelligence research. *ArXiv Preprint ArXiv:1903.00742*. (2019)

[62] Moulin-Frier, C., 2022. The ecology of open-ended skill acquisition. (Professoral thesis, Université de Bordeaux)

[63] Solé, R., 2016. Synthetic transitions: towards a new synthesis. Philosophical Transactions of the Royal Society B 371, 20150438.

[64] Wang, J., Kurth-Nelson, Z., Soyer, H., Leibo, J., Tirumala, D., Munos, R., Blundell, C., Kumaran, D. & Botvinick, M., 2016. Learning to reinforcement learn. *Annual Meeting Of The Cognitive Science Society*.

[65] Duan, Y., Schulman, J., Chen, X., Bartlett, P., Sutskever, I. & Abbeel, P., 2016. RL^2: Fast Reinforcement Learning via Slow Reinforcement Learning. *ArXiv Preprint ArXiv: Arxiv-1611.02779*.

[66] Pedersen, J. & Risi, S., 2021. Evolving and merging hebbian learning rules: increasing generalization by decreasing the number of rules. *Proceedings Of The Genetic And Evolutionary Computation Conference*. pp. 892-900.

[67] Klyubin, A., Polani, D. & Nehaniv, C., 2005. Empowerment: a universal agent-centric measure of control. *Procs Of The 2005 IEEE Congress On Evolutionary Computation 1 Pp. 128-*.

[68] Irrgang, C., Boers, N., Sonnewald, M., Barnes, E.A., Kadow, C., Staneva, J. and Saynisch-Wagner, J., 2021. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. Nature Machine Intelligence 3(8), 667-674.

[69] Qi, D. and Majda, A.J., 2020. Using machine learning to predict extreme events in complex systems. Proceedings of the National Academy of Sciences 117(1), 52-59.

[70] Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A. and Luccioni, A.S., 2022. Tackling climate change with machine learning. ACM Computing Surveys 55(2), 1-96.

[71] Kaack, L.H., Donti, P.L., Strubell, E., Kamiya, G., Creutzig, F. and Rolnick, D., 2022. Aligning artificial intelligence with climate change mitigation. Nature Climate Change 12(6), 518-527.

[72] Lenton, T.M., Held, H., Kriegler, E., Hall, J.W., Lucht, W., Rahmstorf, S. and Schellnhuber, H.J., 2008. Tipping elements in the Earth's climate system. Proceedings of the national Academy of Sciences 105, 1786-1793.

[73] Dakos, V., Matthews, B., Hendry, A.P., Levine, J., Loeuille, N., Norberg, J., Nosil, P., Scheffer, M. and De Meester, L., 2019. Ecosystem tipping points in an evolving world. Nature ecology & evolution 3(3), 355-362.

[74] Scheffer, M., 2020. *Critical transitions in nature and society*. Princeton University Press.

[75] Kéfi, S., Saade, C., Berlow, E.L., Cabral, J.S. and Fronhofer, E.A., 2022. Scaling up our understanding of tipping points. Philosophical Transactions of the Royal Society B 377(1857), 20210386.

[76] Leibo, J., Zambaldi, V., Lanctot, M., Marecki, J. & Graepel, T. Multi-Agent Reinforcement Learning in Sequential Social Dilemmas. *Proceedings Of The 16th Conference On Autonomous Agents And MultiAgent Systems*. 464-473 (2017,2)

[77] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine,. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: *Proceedings of the International Conference on Machine Learning*, 1861–1870.

[78] Karafyllidis, I. and Thanailakis, A., 1997. A model for predicting forest fire spreading using cellular automata. Ecological Modelling, 99(1), 87-97.

[79] Encinas, A.H., Encinas, L.H., White, S.H., Del Rey, A.M. and Sánchez, G.R., 2007. Simulation of forest fire fronts using cellular automata. Advances in Engineering Software, 38(6), pp.372-378.

[80] Alexandridis, A., Russo, L., Vakalis, D., Bafas, G.V. and Siettos, C.I., 2011. Wildland fire spread modelling using cellular automata: evolution in large-scale spatially heterogeneous environments under fire suppression tactics. International Journal of Wildland Fire, 20(5), pp.633-647.

[81] Xu, Y., Li, D., Ma, H., Lin, R. and Zhang, F., 2022. Modeling forest fire spread using machine learning-based cellular automata in a GIS environment. Forests, 13(12), p.1974.