Selfish Algorithm and Emergence of Collective Intelligence

Korosh Mahmoodi^{1*}, Bruce J. West², and Cleotilde Gonzalez¹

1) Dynamic Decision Making Laboratory, Department of Social and Decision Sciences,

Carnegie Mellon University, Pittsburgh, PA 15213, USA

2) Information Sciences Directorate, Army Research Office, Research Triangle Park, NC 27708, USA

We propose a model for demonstrating spontaneous emergence of collective intelligent behavior (i.e., adaptation and resilience of a social system) from selfish individual agents. Agents' behavior is modeled using our proposed selfish algorithm (SA) with three learning mechanisms: reinforced learning (SAL), trust (SAT) and connection (SAC). Each of these mechanisms provides a distinctly different way an agent can increase the individual benefit accrued through playing the prisoner's dilemma game (PDG) with other agents. SAL generates adaptive reciprocity between the agents with a level of mutual cooperation that depends on the temptation of the individuals to cheat. Adding SAT or SAC to SAL improves the adaptive reciprocity between selfish agents, raising the level of mutual cooperation. Importantly, the mechanisms in the SA are self-tuned by the internal dynamics that depend only on the change in the agent's own payoffs. This is in contrast to any pre-established reciprocity mechanism (e.g., predefined connections among agents) or awareness of the behavior or payoffs of other agents. Also, we study adaptation and resilience of the social systems utilizing SA by turning some of the agents to zealots to show that agents reconstruct the reciprocity structure in such a way to eliminate the zealots from getting advantage of a cooperative environment. The implications and applications of the SA are discussed.

Keywords: Selfish Algorithm, Reinforcement Learning, Emergence, Collective Intelligence, Adaptive Reciprocity, Resilience

I. INTRODUCTION

It is nearly two centuries since Lloyd [1] introduced the *tragedy of the commons* scenario to highlight the limitations of human rationality in a social context. In a scenario in which individuals must share a common resource, individuals acting selfishly, without regard for the others sharing the resource, will result in the depletion of resources and subsequent tragedy. But, one can find a number of weaknesses in this arcane argument, which we pose here as questions. Do people always (or ever) behave strictly rationally? Do individuals ever act independently of what others in their community are doing? Do people typically disregard the effects of their actions on the behavior of others?

Reciprocity, the act of behaving towards others in ways that would result in mutual benefit, has been a common answer to the questions posed above. To explain why people behave rationally and whether people care about the effects that their actions would have on a common good, researchers have used predefined norms of reciprocity between agents, such as network reciprocity [4–9], reputation [10, 11], social diversity [12], aspiration [13], punishment [14] and reward [15]. This research has resulted in significant advancements in understanding the emergence of collective behavior, however there are also limitations on the assumptions of predefined reciprocity. First, reciprocity often relies on having explicit knowledge of the decisions others make along with their payoffs; second, reciprocity often assumes the existence of a network structure where agents are linked to each other in some way.

In contrast to much of past research, we address the problem of collective intelligence as an emergent phenomenon, with a model that does not assume any predefined reciprocity. In our proposed algorithm (Selfish Algorithm, SA), agents learn about their dynamic social environment by tracking changes in their own payoffs in response to changes in society and adjust their actions accordingly. We show that reciprocity emerges between agents' mutual interactions and leads them toward stable cooperative behaviour in a Prisoner's Dilemma game (PDG). We also demonstrate that network reciprocity emerges spontaneously, from the selfish mechanisms of the model. Finally, we demonstrate the resilience of this model to changes in the environment. When zealots are introduced, cooperative behavior re-emerges naturally.

In what follows, we focus on the related literature, we then summarize our contributions and the significance of this study, before introducing SA, its components and results.

II. RELATED LITERATURE

Altruism is one concept that was missing from Lloyd's tragedy of the commons discussion, which seems to us, no less important than the notion of selfishness. The latter was foundational to his argument. Moreover, if selfishness is entailed by rationality in decision making, as he maintained, then altruism must be realized through an irrational mechanism in decision making. In point of fact, the competition between the two aspects of human decision making [16, 17], the rational and irrational, may

^{*}Corresponding author: koroshm@andrew.cmu.edu

well save the commons from ruin.

A. The Altruism Paradox

The altruism paradox (AP) identifies a selfcontradictory condition regarding the survival of of a species. This dates back to Darwin's recognition that some individuals in a number of species act in a manner that although helpful to other members of the species, may jeopardize their own survival. He also identified such altruism as contradicting his theory of evolution, that being natural selection [18]. Darwin proposed a resolution to this problem by speculating that natural selection is not restricted to the lowest element of the social group, the individual, but can occur at all levels of a biological hierarchy, which constitutes multilevel selection theory [19].

Rand and Nowak [21] emphasize that natural selection suppresses the development of cooperation unless it is balanced by specific counter mechanisms. Five such mechanism found in the literature are: multilevel selection, spatial selection, direct reciprocity, indirect reciprocity and kin selection. In their paper they discuss models of these mechanisms, as well as, the empirical evidence supporting their existence in situations where people cooperated, but in the context of evolutionary game theory.

In this work, and in agreement with sociobiology research [19, 20], we show how Darwin's theory of biological evolution is compatible with sociology. Cooperation can emerge in the presence of selfish individuals that make up the society.

B. Prisoner's Dilemma and emergence of cooperation

We use the PDG in demonstrations of the emergence of cooperation among a large number of agents. The PDG dates back to the early development of game theory [22], and is a familiar mathematical formulation of the essential elements of many social situations involving cooperative behavior. PDGs are generally represented with a payoff matrix that provides payoffs according to the actions of two players (see Table I). When both players cooperate, each of them gains the payoff R, and when both players defect, each of them gains P. If in a social group agents i and j play the PDG, when i defects and j cooperates, i gains the payoff T and j gains the payoff S and when the decisions are switched, so too are the payoffs. The constraints on the payoffs in the PDG are T > R > P > S and S + T < 2R. The temptation to defect is established by setting the condition T > R.

The dilemma arises from the fact that although it is clear that for a social group (in the short-term) and for an individual (in the long-term) the optimal mutual action is for both to cooperate, each individual is tempted TABLE I: The general payoffs of PDG. The first value of each pair is the payoff of agent i and the second value is the payoff of the agent j.

$$\begin{array}{c} \text{Player } j \\ C \quad D \\ \text{Player } i \begin{array}{c} C \end{array} \begin{bmatrix} (R,R) & (S,T) \\ D & (T,S) & (P,P) \\ \end{array} \end{array}$$

to defect because that decision elicits the higher immediate reward to the individual defecting. But, assuming the other player also acts to selfishly maximize her own benefit, the pair will end up in a defector-defector situation, having the minimum payoff P for both players. How do individuals realize that cooperation is mutually beneficial in the long-term?

This question has been answered by many researchers, at various levels of inquiry, involving pairs of agents [3, 23], as well as, larger social networks [2]. Research suggests that, at the pair level, people dynamically adjust their behavior according to their observations of each others' actions and outcomes: at the level of a complex dynamic network or societal level, this same research suggests that the emergence of cooperation may be explained by network reciprocity, whereby individuals play primarily with those agents with whom they are already connected in a network structure. The demonstration of how social networks and structured populations with explicit connections foster cooperation was introduced by Nowak and May [24]. Alternative models based on network reciprocity assume agents in a network play the *PDG* only with those agents to whom they have specific links. Agents act by copying the strategy of the richest neighbor, basing their decisions on the observation of the others' payoffs. Thus, network reciprocity depends on the existence of a network structure (an already predefined set of links among agents) and on the awareness of the behavior and payoffs of interconnected agents.

a. Empirical evidence of emergence of cooperation Past research has supported the conclusion that the survival of cooperators requires the observation of the actions and/or outcomes of others and the existence of predefined connections (links) among agents. Indeed, empirical work suggests that the survival and growth of cooperation within the social group depends on the level of information available to each agent [25]. The less information about other agents that is available, the more difficult it is for cooperative behavior to emerge [22, 25]. On the other hand, other experiments suggest that humans do not consider the payoffs to others when making their decisions, and that a network structure does not influence the final cooperative outcome [26]. In fact, in many aspects of life, we influence others through our choices and the choices of others affect us, but we are not necessarily aware of the exact actions and rewards received by others that have affected us. For example, when a member of society avoids air travel in order to reduce

their carbon footprint, s/he might not be able to observe whether others are reducing their air travel as well, yet they rely on decisions others make, thereby influencing the community as a whole. Thus, it is difficult to explain how behaviors can be self-perpetuating particularly when the source of influence is unknown [25].

b. The Hypothesis These empirical observations when taken together support an important hypothesis of the work presented herein. Our hypothesis is that *mu*tual cooperation emerges and survives, even when the social group consists exclusively of selfish agents and there is no conscious awareness of the payoffs or decisions of other agents. Moreover an adaptive network structure can emerge dynamically from the connections formed and guided by individual selfishness.

Note that we address the dynamics \underline{of} a network, as distinct from the more familiar dynamics \underline{on} a network. The dynamics on a network assumes a static network structure, as in evolutionary game theory, whereupon the strengths of the links between agents may change, but the agents sit at the nodes of the network and interact with the same nearest neighbors. On the other hand, the dynamics of a network makes no such assumption and the dynamics consist of the formation and dissolution of links between any two agents within the social group. Thus, collective cooperation emerges through establishing of a network that develops over time.

III. CONTRIBUTIONS AND SIGNIFICANCE OF THIS STUDY

Herein, we aim to clarify how collective intelligence emerges without explicit knowledge of the actions and outcomes of others, and in the absence of any predefined reciprocity between the agents (such as network reciprocity linking agents within a society). We introduce an algorithm (the *Selfish Algorithm*, *SA*) and construct a computational model to demonstrate that collective intelligence can emerge and survive between agents in a social group, out of selfishness.

First, the SA model provides a resolution of the altruism paradox (AP) and shows how agents create a selforganized critical group out of individual selfish behavior, simultaneously maximizing the benefit of individual and of the whole group. Second, the SA model demonstrates how *adaptation* can naturally emerge from the same mechanisms in the model. Adaptation is an important property of living things, which allows them to respond to a changing environment in a way that optimizes their performance and survivability.

We studied four systems governed with different learning mechanisms proposed by the SA, where the behavior of the agents is modeled using three modes of learning: reinforced learning (SAL), trust (SAT), and connection (SAC). The first system studied has only the SAL mechanism active. The other three systems are a combination of SAL and SAT (SALT), SAL and SAC (SALC), and the combination of all the mechanisms (*SALTC*). Next, we tested the sensitivity of the collective behavior that emerged from these systems to changes in the social makeup by modifying a fraction of the agents by having them exchanged with zealots at a given time. A zealot is an agent that will not change its decision regardless of the payoff; defecting all the time. We present a comparison of the mutual cooperation of a system with the same sized system having a number of zealots to provide a measure of resilience, or robustness (i.e., the resistance of the collective behavior to perturbations). The computational results show that these systems can be ranked from strongest to weakest as a function of their resilience.

To put the significance of our contributions in perspective, let us highlight the main differences between the results of the SA model and those of previous studies. First, the works following the strategy of the Nowak-May agents [24] typically compare their individual payoff with the payoffs of the other agents when making their next decision. One way to realize this strategy is to adopt the decision of the most successful neighbor. In contrast, all the updates in the SA model are based on the last two payoffs to the self-same individual. In other words, SA agents do not rely on information about the other members of society to determine their evolution. SA agents are connected to other agents only indirectly through their local payoffs in time. The most recent payoff to the individual is the result of its interaction with its most recent partner, but its previous payoff, used for tuning its tendencies (presented below in Equation 1), might be the result of playing with a different partner. Second, the Nowak-May agents [24] are assumed to be located at the nodes of a lattice and to be connected by links, as they were also assumed to be in the decision making model (DMM) [20]. For example, because of social interactions these lattice networks are regularly used to model complex networks and have historically been assumed to mimic the behavior of real social systems. These studies emphasized the important role of network complexity in sustaining, or promoting, cooperation [28]. This was accomplished without explicitly introducing a self-organizing mechanism for the network formation itself.

In appendix A we replicate the simulation work of Nowak and May [24] to demonstrate the limitations of having predefined network reciprocity mechanisms. First, we showed that in their model mutual cooperation only exists for small values of temptation to cheat (T_c) and for high values of initial cooperators on the lattice (see Figure 12). Second, we show that by removing the predefined network reciprocity between the agents the mutual cooperation cannot survive. (see Figure 13).

As we demonstrate next, the SA model does not rely on a predefined network reciprocity, yet, mutual cooperation is subsequently shown to emerge and survive the disruption caused by perturbations. In fact, we show that the network reciprocity can be interpreted to be a byproduct of the SA, emerging out of the selfishness of

IV. THE SELFISH ALGORITHM (SA)

The selfish algorithm (SA) proposed in this research belongs to a family of agent-based models in modern evolutionary game theory [29]. Specifically, the SAmodel represents a generalization of a model introduced by Mahmoodi et al. [8] which is a coupling of DMMwith evolutionary game theory put into a social context. This earlier introduction led to the self-organized temporal criticality (SOTC) concept and the notion that a social group could and would adjust their behavior to achieve consensus, interpreted as a form of group intelligence, as had been suggested by the collective behavior observed in animal studies [30]. The SOTC model, like the Nowak and May model [24], assumes a pre-existing network between the agents which use the DMM (belonging to the Ising universality class of models for decision making [31]) to update their decision. In Appendix B, we give a summary of the DMM and SOTC. The SAmodel overcomes these constraints, resulting in a general model for emergence of collective intelligence from a complex dynamic network.

The general notion that the SA model adopted from Mahmoodi et al. [8] is that an agent *i* makes a decision according to the change in a cumulative tendency (i.e., preference). The change in this cumulative tendency, Δ , is a function of the last two payoffs of agent *i* who played with agents *j* and *k*:

$$\Delta_{i,jk} = \chi \frac{\Pi_{ij}(t) - \Pi_{ik}(t-1)}{|\Pi_{ij}(t)| + |\Pi_{ik}(t-1)|},$$
(1)

where χ is a positive number that represents the sensitivity of the agent to its last two payoffs. The quantity $\Pi_{ij}(t)$ is the payoff agent *i* obtained from the play with agent *j* at time *t* and $\Pi_{ik}(t-1)$ is the payoff to the agent *i* obtained from the play with agent *k* at time t-1. We used the *S* value in the *PDG* that is > 0 and assumed $\Delta_{i,jk} = 0$ when the denominator of Eq. (1) is zero.

A simplified version of the SA model was previously introduced elsewhere [32], and a detailed version of the SA algorithm and its mathematical details are included in Appendix C.

In the first cycle of the SA, a pair of randomly selected agents *i* and *j* "agree" to play. Only one pair of agents play at each time cycle. Each agent of a pair engages in three decisions; each decision is represented in a learning mechanism (a cumulative propensity lever, ranging from 0 to 1) by which the agent can gain information about the social environment: learning from their own past outcomes (Selfish Algorithm Learning, SAL), trust the decision of other agents (Selfish Algorithm Trust, SAT), and make social connections that are beneficial (Selfish algorithm-based connection, SAC). Each of these three levers is updated according to the agent's self interest (selfishness): a decision is made according to the state of each lever, and the lever is updated according to Δ as formalized in Eq. (1).

In the SAL mechanism, an agent decides to cooperate (C) or defect (D) while playing the PDG with the paired partner, according to the state of the cumulative propensity of playing C or D. The cumulative propensity of the agent to pick C or D increases (or decreases) if it's payoff is increased (or decreased) with respect to its previous payoff, as per Eq. (1). The updated cumulative tendency to play C or D with the same agent is used for the next time agent i is paired with the same agent j. The simulation results show that the SAL mechanism attracts the agents toward mutual cooperation.

In the SAT decision, an agent decides to rely on the decision of its partner, instead of using its own decision made using the SAL model, according to the state of the cumulative propensity of "trusting" the other's decision or not. Each SAT agent can tune this propensity to rely on its partner's decision according to Δ as formalized in Eq. (1). The SAT agent increases (or decreases) its tendency to trust its partner's decision if it increased (or decreased) its payoff with respect to its previous payoffs. The simulation results show that the SALT (both SAL and SAT active) mechanism amplifies the mutual cooperation between the agents regardless of the value of the incentive to cheat T_c .

Finally, the SAC model determines the decision of an agent to pick the partner with whom to play in each round. Each SAC agent can tune its propensity of selecting its partner according to Δ as formalized in Eq. (1). A SAC agent increases (or decreases) its tendency of playing with the same partner if the payoff received after playing with that partner is higher (or lower) than the agent's own previous payoff. The simulation results show that a network of connections emerge over time between SAC agents, and that this network amplifies the mutual cooperation between agents.

The details of the updating rules and evaluation of the propensities of the decisions are presented in Appendix C. As an example, here we explain the updating of the cumulative tendencies and the evaluation of the propensities for decisions C or D (i.e., the SAL mechanism), which presents the generic way of learning in the SA: After paired agents i and j received their payoffs they update their cumulative tendencies for the decisions they made. The cumulative tendencies for decisions C or D update as follows: If the decision of agent i was C then the cumulative tendencies C_{ij} and D_{ij} change to $C_{ij} + \Delta_{i,jk}$ and $D_{ij} - \Delta_{i,jk}$ for the next time agent i and j paired. If the decision of agent i was D then the cumulative tendencies D_{ij} and C_{ij} change to $D_{ij} + \Delta_{i,jk}$ and $C_{ij} - \Delta_{i,jk}$ for the next time agent *i* and *j* paired. The same happens for agent j, with the indices interchanged. The propensity that agent i picks decision Cor D, if is paired with agent j, is simply $C_{ij}/(C_{ij}+D_{ij})$ or $D_{ij}/(C_{ij} + D_{ij})$, respectively. The same happens for agent j, with the indices interchanged. The other tendencies are updated in a similar way.

TABLE II: The payoffs of the PDG. The first value of each pair is the payoff of agent i and the second value is the payoff of the agent j.



V. SIMULATION METHODS

We fixed the number of agents to 20 in all the simulations. However, for robustness, we replicate the results presented below with 10, 30, 40, and 50 agents in Appendix D. The results indicate that the size of the system alters the learning time of the agents but does not affect the emergence of cooperation in the long run.

We set the following initial parameters for the calculations: all the agents started as defectors, had payoffs of zero, propensities of 0.9 to remain defectors, propensities of zero to trust their partner's decision, with equal propensities of 1/(20-1) to connect with each of the other agents, and the sensitivity coefficient $\chi = 200$. The parameter χ controls the magnitude of the changes of the cumulative tendencies. Again, for testing the robustness of the model, we replicated the simulations with other value of $\chi = 400$ which are presented in Appendix D. The results show that a higher χ , with respect to the maximum value of the cumulative tendency, improves the process of learning, but it does not qualitatively change behavior, but merely scales the dynamical properties of the system.

The payoff matrix used in the simulations is shown in Table II as suggested by Gintis [33]: R = 1, P = 0 and S = 0. So, the maximum possible value of $T = 1 + T_c$ is 2. We selected the value $T_c = 0.9$ (unless otherwise explicitly mentioned) which provides a strong incentive to defect. This simple payoff matrix has the spirit of the PDG and allows us to emphasize the main properties of the SA model. However, for completeness and for testing the robustness of our model we also ran simulations with the constraints of T > R > P > S by making P range between 0.0 and 0.2. The results are presented in Appendix D and show that SA can generate cooperation regardless of increase of P, although, in a lowered amount.

VI. SIMULATION RESULTS

A. Emergence of reciprocity and mutual cooperation by *SAL*

In this section we demonstrate that the learning mechanism of SAL model attracts agents playing the PDGto a state of mutual cooperation. Figure 1 shows the time evolution of the RMC in simulations where the 20 agents, randomly partnered in each cycle, used only the



FIG. 1: Each curve shows the time evolution of the Ratio of Mutual Cooperation (time average of mutual cooperation at time t) for 20 agents used *SAL* and

played the PDG with $T_c = 0.1, 0.3, 0.5, 0.7, 0.9$ corresponding to the black, red, blue, green and purple curves, respectively. The standard error of the curves (for $5 \times 10^5 < t < 10^6$) is about 1.95×10^{-6} . Each curve is the result of a single simulation.

SAL mechanism to update their decisions. The RMC increased from zero and asymptotically saturates to a value that depends on the value of the temptation to cheat T_c . The smaller the T_c value the higher the saturated RMC value and the greater the size of the collective cooperation. Using SAL, each agent learns, by social interactions with the other agents, to modify its tendency to play C or D. These agents are connected through their payoffs. Each agent compares its recent payoff with its previous one, which, with high propensity, was earned by playing with a different agent. This mutual interaction between the selfish agents led them to form an intelligent group that learns the advantage of mutual cooperation in a freely connected environment whose dynamics are represented by the PDG.

To explain how reciprocity emerges, we looked into the evolution of the propensities to cooperate between two selected agents (among 20). Figure 2 depicts the time evolution of the propensity of agent 1 to cooperate with agent 2 ((PC_{12}) , red curve), and the propensity of agent 2 to cooperate with agent 1 ((PC_{21}) , black curve). Generally, we observe a high correlation between the propensities to cooperate between the two agents. The top panel presents these propensities when the temptation to defect is $T_c = 0.9$, and the bottom panel presents the propensities when the temptation to defect is $T_c = 0.1$. The correlation between the propensity of the two agents is 0.65 for $T_c = 0.9$ and 0.84 for $T_c = 0.1$. The correlation between the agents is a consequence of the learning process between them. Let's assume $T_c = 0.5$ and that agent 1 played C and agent 2 played D at time t which results



FIG. 2: Emergence of decision reciprocity: The curves show the time evolution of the propensity of agent 1 to

play as C with agent 2 (red curve) (PC_{12}) and the propensity of agent 2 to play as C with agent 1 (black curve) (PC_{21}) . Agents 1 and 2 were among 20 agents used SAL to update their decisions and had $T_c = 0.9$ (top panel) and $T_c = 0.1$ (bottom panel). The standard error of the curves of the top and bottom panels are

about 4.71×10^{-4} and 3.45×10^{-4} , respectively. The curves in each panel are the result of a single simulation.

in payoffs of 0 and 1.5 for them, respectively. Comparing its payoff with its previous payoff, earned playing with other agent (= 1.5, 0, 1 or 0), agent 1 would change its accumulative tendency to play C for the next time it is randomly paired with player 2 by $-\chi$, 0, $-\chi$ or 0. This means that agent 1 reacts to the defective behavior of agent 2 by tending to behave as D and consequently agent 2 wouldn't continue to have the advantage of a cooperative environment. On the other hand, agent 2 would change its accumulative tendency to play D with agent 1 by 0, χ , 0.2 χ or χ . This means agent 2 would like to play as D next time it pairs with agent 1. So, both agents learn to play D with one another, leading to

the coordination state DD where both get a zero payoff. Such pairs compare this payoff (= 0) with their previous payoff, played with other agents, and would change their accumulative tendency to play D by $-\chi$, 0, $-\chi$ or 0, which shifts their future decisions toward the coordination state CC. These agents would change their tendency to play C by -0.2χ , χ , 0 or χ which favors their stay as C toward one another. However, because of the time to time change of -0.2χ (depending on the value of T_c) in the accumulative tendency to play C with other pairs, there is always a chance for agents to play D for a while, before they are pulled back by other agents to behave as C. This creates a dynamic equilibrium between CC and DD states and defines the level of emerged mutual cooperation observed in Figure 1. Thus, the dual dynamic interaction, based on self-interest, between each agent and its environment causes the emergence of mutual cooperation between the agents who play PDG and use the SAL model to update their decisions. Note that in human experiments, humans learning from only their own outcomes without awareness of the partners' outcomes did not lead to mutual cooperation [25] as distinct from what the rational agents of the SAL model do. High correlation between the propensities of the pairs of agents using the SAL model means high coordination between their decisions, which also can occur if the "Trust" mechanism is active between the agents. Trust lowers the intermediate CD pairings and leads agents to coordinate and converge on the same decision. In the next section we show that combining the SAL model with the SATmodel amplifies the RMC between the agents. Our current experimental work [34] also confirms that adding the trust mechanism in decision making experiments with of human pairs assisted them in realizing the advantage of mutual cooperation.

B. Enhancement of mutual cooperation by *SALT*; trust as an adaptive decision

Figure 3 shows the enhancing effect on the emergence of collective cooperation when each agent is allowed to make a decision whether to "Trust" (rely on) the decision made by the paired agent or not (SAT).

The SALT mechanism also decreases the time for agents to realize the benefit of mutual cooperation. In Figure 4 we compare the CD and DD ratios of partners in a group of 20 agents playing the PDG with a temptation to cheat of $T_c = 0.9$ when using the SALT model (dashed lines) and when using the SAL model (solid lines). It is apparent that because of the trust mechanism, SALT agents coordinate more often than they do when the trust mechanism is absent and consequently avoid the formation of CD pairs. We also plot the ratio of DD pairs for the two systems. These curves show that SALT agents learn to select CC pairs over CDpairs more readily than do SAL agents, thereby pushing the RMC up to approximately 0.9. The average correla-



FIG. 3: Each curve shows the time evolution of the Ratio of Mutual Cooperation (RMC) for 20 agents used SALT and played the PDG with $T_c = 0.1, 0.3, 0.5, 0.7, 0.9$ corresponding to the black,

red, blue, green, and purple curves, respectively. Each curve is the result of a single simulation.



FIG. 4: The solid black and the solid red curves show, respectively, the time evolution of the ratio of CD and DD pairs played PDG and used SAL to make decision. The dashed black and dashed red curves show,

respectively, the time evolution of the ratio of CD and DD pairs played PDG and used SALT to make

decision. M = 20, $T_c = 0.9$. Each curve is the result of a single simulation.

tion between the propensity of two agents to play C with one another is about 0.92 when $T_c = 0.9$ which is a sign of high coordination. We highlighted the difference between trust used in the literature with our dynamic trust model in Appendix E.

C. Emergence of network reciprocity from SALCand SALTC

By activating the ability of the agents to make decisions about the social connections that are beneficial to themselves (SAC), we expect that an even larger and faster increase in cooperative behavior. The connection mechanism allows each agent to select a partner that helped the agent to increase its payoff with respect to its previous payoff.

We demonstrate the increase in the RMC for a model without the Trust mechanism SALC and a model with the Trust mechanism SALTC. The top panel of Figure 5 shows an increase in the RMC of the agents using SALC (top panel), and using SALTC (bottom panel). When comparing the SALC model behavior to that of the agents without the connection mechanism (paired randomly) in Figure 1, we observe that the dependence on the temptation to cheat T_c is weaker. For example at time $t = 10^6$ the average RMC for the agents using SALC are about 0.84, 0.80, 0.75, 0.65 and 0.62 for $T_c = 0.1, 0.3, 0.5, 0.7$ and 0.9, respectively, whereas for the agents using SAL the average RMC for the same conditions are about 0.76, 0.64, 0.58, 0.52 and 0.45, respectively. On the bottom panel of Figure 5, we observe that using the Learning, Trust, and Connection mechanisms in conjunction (SALTC), the level of cooperative behavior is the highest with the least dependence on the temptation to cheat.

To show that the reciprocity emerged between the agents using the SAC mechanism, we plotted in Figure 6 the propensities of making connections between a typical agent (agent 1) and the other 19 agents where agents used SALC (top panel) or SALTC (bottom panel) to update their decisions. In both cases, these figures show that agent 1 developed a preferential partner and learned to play most of the time with one of the agents among others.

The manner in which the network develops over time is schematically depicted in Figure 7. Three panels of this figure show the snapshots of the connection propensities between 20 SA model agents taken at three times. Intensity of the lines between pairs show the magnitude of the propensity of one agent to connect to the other and the directions indicates that the intensity of the affinity one agent has for another is asymmetric. Agent A can have a high affinity for agent B, but agent B does not necessarily reciprocate with the same level of affinity for agent A. The colors of the nodes represent the state of the agents at that time, red as defector and green as cooperator. The figure shows the connections among the agents at $t = 10^1$ (left panel), passing through an intermediate state (middle panel, $t = 5 \times 10^3$) and after reaching dynamic equilibrium (right panel, $t = 5 \times 10^5$). The preferential connections forming the dynamic network emerge here over time and are based on the perception of the benefit that an agent receives from other agents, with whom it has interacted. Some connections become stronger,





corresponding to the black, red, blue, green and purple curves, respectively. Each curve is the result of a single simulation.

whereas others become weaker according to the SAC mechanism. The IPL PDF is very different from the Poisson PDF, the latter having an exponentially diminishing probability of changing the propensity compared with the much greater IPL value. Consequently, the propensity of forming a link between partners is much greater for the IPL PDF and the SAC model over time forms a much more complex network than does a Poisson PDF.

In the next section we study the adaptability of social systems ruled by different size steps in the SAmodel. Disrupting these systems is done by changing some agents into zealots and tracking the changes among the remaining agents in response to the zealots.



FIG. 6: Top and bottom panels show the time evolution of the propensity of pairings between agent 1 and the other 19 agents when agents used SALC or SALTC, respectively. Agents had $T_c = 0.9$. The curves in each panel are the result of a single simulation.

D. SA entails complex adaptive reciprocity

To investigate the dynamics of the SA agents we disrupt the stable behavior pattern emerging from the social group by fixing a fraction f = 0.5 of the N = 20agents to be zealots and calculating the response of the remaining agents. A zealot is an agent whose properties are: zero tendency to be a C; zero trust to other agents; and a uniform tendency to play the PDG with other agents. This choice divides the system into two subsystems: a fraction f = 0.5 of the agents that continue to evolve based on SAL, SALT, SALC or SALTC (subsystem S) and (1 - f)N agents as zealots (subsystem \overline{S}). It would appear that the zealots might have the advantage of defection in the emerging cooperative environment. We investigate how the remaining fN agents of system S adapt to the new environment by various modes of learning. The degree to which these agents can sustain their mutual cooperation in the presence of the



FIG. 7: Emergence of reciprocity from dynamics of the network. From left to right the panels show the snapshots of the propensities of connections between 20 agents played the *PDG* and updated their decisions based on *SALTC* at $t = 1, t = 5 \times 10^3$ and $t = 5 \times 10^5$, respectively. The Intensity of the lines between nodes represent the magnitude of the propensity of one agent to select the other to play PDG. The direction of the lines show the intensity belongs to which agent and towards which one. The colors of the nodes represent the state of the agents at the times the

snapshots were taken

, red as defector and green as cooperator.

zealots is a measure of their resilience, or its compliment is a measure of their fragility, a fundamental property of complex networks subject to perturbation.

To track the behavioral changes of the agents in \overline{S} from those in S we study the chance of an event happening within a given SA cycle. This could be the chance of finding the pairings Cooperation-Cooperation, Trust-Trust, etc. In previous sections we used the ratio of events, which was useful as there was no perturbation in the systems to detect. To evaluate the chance of the event occurring we used ensemble averages over 10^3 simulations of each simulation.

The blue and orange curves in the panels of Figure 8 show the *CMC* within a group of N = 20 agents and CMC within the subsystem of the 10 agents (CMC_S) who were not exchanged with zealots after time t_z . The top-left panel, shows the CMC and CMC_S between the agents which used SAL to update their decisions, before and after 10 of them being replaced by zealots at time $t_z = 2.5 \times 10^5$. There is a drop in the *CMC* because of the inevitable pairings between the SAL model agents and the zealots. The CMC_S between the 10 agents who were not switched with zealots increased after their interactions with the zealots, despite the overall decrease in the CMC in the system. In other words, the agents of the subgroup S improved their mutual cooperation from about 0.1 to about 0.25. This is because when the agent of subsystem S is randomly paired with a zealot of the subsystem \overline{S} , with high probability, it ended up as a sucker and received the minimum payoff of zero. The agent used this payoff as a measure on which to base its next decision. When an agent of subsystem S paired with another agent of the same subsystem, with high probability (because of their past experience of playing together) played C, but because of the low payoff it received previously playing with a zealot, still increases its tendency to play C with this agent.

The top-right panel in Figure 8 depicts the CMC and

 CMC_S between agents who used the SALT model to make decisions, before and after the switching time $t_z =$ 2.5×10^5 . Although the SALT model highly increased the CMC level, it failed to sustain that level due to the influence of the zealots. After the switching time t_z only the agents of the subsystem S contributed to the mutual cooperation.

The bottom-left panel in the figure depicts the advantage of adding the SAC to the SAL mechanism in sustaining the CMC of the system after the switching time $t_z = 2.5 \times 10^6$. This figure shows that after the time t_z the CMC_S of the subsystem S increases, as the only contribution for mutual cooperation, and saturates at about 0.85. The bottom-right panel of the figure shows the CMC and CMC_S of the agents use SALTC (entire SAalgorithm) to update their decisions. The panel shows a very high resilience of the system even after this massive number of SA agents switched to zealots at time $t_z = 2.5 \times 10^6$. Similar to the system governed by the SALC model, there is a drop in the CMC but the agents could sustain the level of CMC_S to about 0.95. This is because the SA agents of this system learned to disconnect themselves from the zealots using the SAC mechanism, which highly increased the robustness of the emerging mutual cooperation. Notice that after the switching time t_z all the mutual cooperation occurs within subsystem S.

Figure 9 summarizes the influence a given fraction of zealots within a group manifest under the different learning mechanisms. The figure shows the saturated value of CMC_S of the SA agents, using different mechanisms of the SA algorithm for decision making, after a fraction f of the agents turned to zealots at time t_z . The more adaptable the system, the higher the saturation value of CMC_S that can be achieved by the remaining (1 - f)Nagents. This provides us with a measure of the resilience of the system. The solid curve is the saturated CMC_S for the agents using the SAL model as a function of the



FIG. 8: The blue curves show the time evolution of the Chance of Mutual Cooperation (CMC) (chance of CC to happen among 20 agents at time t) and the orange curves shows the CMC between 10 agents who didn't forced to be zealot. In top-left panel agents used SAL and in top-right panel used SALT to update their decisions while, for both cases, since $t_z = 2.5 \times 10^5$ 10 of the agents forced to be zealots. In bottom-left panel used SALTC to update their decisions while, for and in bottom-right panel used SALTC to update their decisions while, for both cases, since $t_z = 2.5 \times 10^6$ 10 of the agents forced to be zealots. N = 20, $T_c = 0.9$. The curves are ensemble averages over 10^3 simulations.

fraction (f) of the agents who switched to zealots at time t_z . For f < 0.4 (8 zealots) the system retains a CMC_S slightly above 0.4. Beyond this fraction of zealots there is an exponential decay of CMC_S . In this situation the SA agents learn to modify their decisions (C or D) depending on whether or not their pair is a zealot or another SA agent using SAL to make its decisions.

The dashed curve in the figure is the saturated value of the CMC of the SA agents, using the SALT model, after the switching time t_z . Introducing the SAT learning improves the resilience of the system for f < 0.4 above that of SAL alone. However, beyond f = 0.4 the calculation converges with the earlier one indicating that the additional learning mechanism of trust ceases to be of value beyond a specific level of zealotry.

The two top curves of Figure 9 are saturated CMC for the SA agents using the SALC model (doted curve) and the SALTC model (dot-dash curve) to make decisions, after the switching time t_z . These two curves show substantial improvement in the system's resilience to an increase in the fraction of zealots. Or said differently, the robustness of the system to perturbations is insensitive to the number of zealots, that is, until the zealots

constitute the majority of group membership. The SAC learning in the decision making of the SA agents have the ability to avoid pairing with the zealots. Once the fraction of zealots exceeds 0.6 the resilience drops rapidly. The saturated CMC for the SALTC model is highest where the three learning levels (decision, trust, connection) are active, giving the SA agents maximum adaptability, that is, flexibility to change in order to maximize their self-interest. This realization of self-interest improves the payoff of the whole system as well. Notice that as the number of the zealots increases, it takes longer for agents to agree to play, but when they do play, they do so with a high propensity that they will cooperate with each other. In other words, increasing the number of zealots slows down the response of the system.

VII. EFFECT OF ZEALOTS ON THE DYNAMICS OF THE SA

Figure 10 shows the time evolution of the propensity of the SAC model of one agent 1 to the other 19 agents, in a system of 20 agents which used the SALT model (top



FIG. 9: Effect of ratio of the zealots (agents that turned to defectors at $t_z = 2.5 \times 10^6$) on the saturated value of the Chance of Mutual Cooperation (*CMC*) of the

remained agents. The agents used SAL (solid curve), SALT (dashed curve), SALC (dot curve) and SALTC(dot-dash curve) to update their decisions. N = 20, $T_c = 0.9$. The curves are ensemble averages over 10^4 simulations.

panel) or the *SALTC* model (bottom panel) to make decision until t_z where 10 (out of the 19) agents turned to zealots. This figure shows that agent 1, before $t = t_z$, mostly played with the agent corresponding to the purple or green curve of top or bottom panel respectively. But after these agents switched to zealots then agent 1 decreased its propensity to pair up and play with them and switched to pairing with the agent corresponding to blue curves (not a zealot).

The three panels of Figure 11 show the snapshots of the propensities of connections between 20 agents, used the *SALTC* model to make decisions, at the beginning state (left panel), at $t = 5 \times 10^5$ where the dynamic network formed and agents learned to do mutual cooperation for their benefit (middle panel) and at $t = 10^6$ where agent 16 had been behaving as a zealot since $t = 5 \times 10^5$ (right panel). This figure shows that the network is intelligent and is able to isolate the zealot, in order to sustain the highest level of mutual cooperation possible.

VIII. DISCUSSION AND IMPLICATIONS OF RESULTS

The Selfish Algorithm provides a demonstration that the benefit to an individual within a social group need not be achieved at the cost of diminishing the overall benefit to the group. Each individual within the group may act out of self-interest, but the collective effect can be quite different than what is obtained by the naive linear logical extrapolation typically made in the tragedy



FIG. 10: Time evolution of the propensity of agent 1 to pair with the other 19 agents $(P_{1j}, j = 2:20)$, where agents used *SALC* (top panel) and *SALTC* (bottom panel) before $t_z = 2.5 \times 10^6$ and after t_z 10 of the agents (out of 19) turned to zealots. N = 20, $T_c = 0.9$. The curves in each panel are the result of a single simulation.

of the commons arguments [1]. In fact, the SA model demonstrates how robust cooperation can emerge out of the selfishness of the individual members of a system that improve the performance of each agent, as well as, that of the overall group, resolving the altruism paradox.

A collective group intelligence emerges from the SA model calculations, one based on the self-interest of the individual members of the group. This collective intelligence grows spontaneously in the absence of explicit awareness on the part of individuals of the outcomes of the behavior of other members of the group. Perhaps more importantly, the collective intelligence develops without assuming a pre-existing network structure to support the collective behavior, and without assuming knowledge of information about the actions or outcomes from other members of the society.

As demonstrated in this research, collective intelligence



FIG. 11: Complex adaptation in the dynamic network emerged by SA. From left to right the panels show the snapshots of the propensities of connections between 20 agents at t = 1, $t = 5 \times 10^5$ and $t = 10^6$, respectively. Agents played the *PDG* and updated their decisions based on *SALTC* except agent 16 (red node on the left panel) which turned to a zealot and kept defecting after $t_z = 5 \times 10^5$. Intensity of the lines between pairs represent the magnitude of the propensity of one to another and the directions show the intensity belongs to which agent and towards which one. The colors of the nodes represent the state of the agents at the times the snapshots were taken , red as defector and green as cooperator. M = 20, $T_c = 0.9$.

entailed by the SA model unfolds as a consequence of three learning mechanisms:

The SAL model supports reinforced learning based on the selfishness of agents who play the PDG. An agent chooses C or D in its play with other agents and agents influence one another through their payoffs, resulting in the emergence of mutual cooperation, that reveals a collective intelligence.

The SAT model tunes the propensity of an agent to imitate the strategy of their partner. The strength of the tuning is proportional to the trust being beneficial or detrimental to the agent itself. The role of the SAT mechanism is to assist agents in achieving coordination, which results in a reduction in the formation of CD pairs. This reduction makes it easier for the system to learn to select CC over CD pairings over time.

The SAC model tunes the propensity of an agent to connect to other agents based on how they improved its self-interest.

We have studied the advantage of each of the SA learning mechanisms in creating mutual cooperation and thereby the facilitation of the growth of group intelligence. We also studied the adaptation and resilience of the emergent intelligence by replacing some of the agents with zealots and examining the system's response. The control of the dynamics in the SA model is internal and according to the self-interest of the agents, spontaneously directs the system to robust self-organization. The most robust systems use the SAC mechanism to make decisions, which increase the complexity of the system by forming a dynamic complex network with a high propensity for partnering, thereby providing the agents with the ability to learn to shun the zealots. These complex temporal pairings that emerge from the SAC model can be seen as a source of the network reciprocity introduce by Nowak and May [24].

One advantage of the SA model is its simplicity, which enables us to introduce additional behavioral mechanisms into the dynamics such as deception or memory. This makes it possible to create algorithms for different selforganizing systems such as those used in voting. The SA model leadership emerges anywhere and everywhere as the need for it arises and it leads to socio-technical systems which achieve optimum leadership activities in organizations to make teams more effective at achieving their goals. The model also has promise in the information and communication technology field of sensor and sensor array design. Future research will explore the application in detecting deception in human - human or human - machine interactions, in anticipating threats, as well as providing leaders with near real-time advice about the emerging dynamics of a given specific situation.

The predictions in this work are entirely simulationbased, however a number of them are being experimentally tested. Experimental work (paper in preparation) confirms the emergence of mutual cooperation between human pairs playing PDG when they are given the option to trust their pair's decision, in no explicit awareness of the payoffs to other agents. Additional experiments to test our hypothesis that the cooperative behavior emerging from the SA model is robust against perturbations are in progress.

Acknowledgments

This work was supported by the Army Research Office, Network Science Program, Award Number: W911NF1710431, to Cleotilde Gonzalez.

- Lloyd, W.F. (1833) Two Lectures on the Checks to Population: Delivered Before the University of Oxford, in Michaelmas Term 1832. JH Parker.
- [2] Nowak, M.A. & Sigmund, K. (1993) A strategy of winstay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, **364**, 56–58.
- [3] Gonzalez, C., Ben-Asher, N., Martin, J.M. & Dutt, V. (2015) A Cognitive Model of Dynamic Cooperation With Varied Interdependency Information. *Cognitive Science*, 39, 457–495.
- [4] Nowak, M.A. (2006) Five rules for the evolution of cooperation. Science, 314, 1560–1563.
- [5] Wang, Z., Kokubo, S., Tanimoto, J., Fukuda, E. & Shigaki, K. (2013) Insight into the so-called spatial reciprocity. *Physical Review E*, 88, 042145.
- [6] Szabó, G. & Tőke, C. (1998) Evolutionary prisoner's dilemma game on a square lattice. *Physical Review E*, 58, 69.
- [7] Masuda, N. & Aihara, K. (2003) Spatial Prisoner's Dilemma optimally played in small-world networks. *Physics Letters A*, **313**, 55–61.
- [8] Mahmoodi, K., West, B.J. & Grigolini, P. (2017) Selforganizing complex networks: individual versus global rules. *Frontiers in physiology*, 8, 478.
- [9] Lee, H.W., Malik, N. & Mucha, P.J. (2018) Evolutionary prisoner's dilemma games coevolving on adaptive networks. *Journal of complex networks*, 6, 1–23.
- [10] Xia, C.Y., Ding, S., Wang, C.J., Wang, J. & Chen, Z.Q. (2017) Risk analysis and enhancement of cooperation yielded by the individual reputation in the spatial public goods game. *IEEE Systems Journal*, **11**, 1516.
- [11] Li, X., Sun, S. & Xia, (2019) Reputation-based adaptive adjustment of link weight among individuals promotes the cooperation in spatial social dilemmas. *Applied Mathematics and Computation.* **361**, 810–20.
- [12] Huang, K.K., Zhao, X.F., Yu, Z.F., Yang, C.H. & Gui, W.H. (2018) Heterogeneous cooperative belief for social dilemma in multi-agent system. *Applied Mathematics and Computation*, **320**, 572–9.
- [13] Geng, Y., Hu, K., Shen, C., Shi, L. & Wang, Z. (2019) Aspiration induced interdependence leads to optimal cooperation level. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **29**, 083114.
- [14] Helbing, D., Szolnoki, A., Perc, M. & Szabó, G. (2010) Punish, but not too hard: how costly punishment spreads in the spatial public goods game. *New Journal of Physics*, **12**, 083005. linear Science, **29**, 083114 (2019).
- [15] Li, X., Wang, H., Xia, C. & Perc, M. (2019) Effects of reciprocal rewarding on the evolution of cooperation in voluntary social dilemmas. *Frontiers in Physics*, 7, 125.
- [16] Ariely, D. (2008) Predictably irrational: The hidden forces that shape our decisions. *HarperCollins*.
- [17] Kahneman, D. (2011) Thinking, fast and slow. *Macmillan*.
- [18] Darwin, C. (1871) The Origin of the Species and the Descent of Man. Modern Library, NY.
- [19] Wilson, D.S. & Wilson, E.O. (2007) Rethinking the theoretical foundation of sociobiology. *The Quarterly Review* of Biology, 4, 327–348.
- [20] West, B.J., Mahmoodi, K. & Grigolini, P. (2019) Empirical Paradox, Complexity Thinking and Generating New

Kinds of Knowledge. Cambridge Scholars Publishing.

- [21] Rand, D.G. & Nowak, M.A. (2013) Human cooperatio. Trends in cognitive sciences, 17, 413–425.
- [22] Rapoport, A. & Chammah, A.M. (1965) Prisoner's Dilemma: A Study in Conflict and Cooperation. University of Michigan Press.
- [23] Moisan, F., ten Brincke, R., Murphy, R.O. & Gonzalez, C. (2018) Not all Prisoner's Dilemma games are equal: Incentives, social preferences, and cooperation. *Decision*, 5, 306–322.
- [24] Nowak, M.A. & May, R.M. (1992) Evolutionary games and spatial chaos. *Nature*, **359**, 826.
- [25] Martin, J.M., Gonzalez, C., Juvina, I. & Lebiere, C. A (2014) Description-Experience Gap in Social Interactions: Information about Interdependence and Its Effects on Cooperation. *Journal of Behavioral Decision Making*, 27, 349–362.
- [26] Fischbacher, U., Gächter, S. & Fehr, E. (2001) Are people conditionally cooperative? Evidence from a public goods experiment. *Economics letters*, **71**, 397–404.
- [27] Barabási, A.L. & Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.
- [28] Tomassini, M., Pestelacci, E. & Luthi, L. (2007) Social dilemmas and cooperation in complex networks. *Interna*tional Journal of Modern Physics C, 18, 1173–1185.
- [29] Adami, C., Schossau, J. & Hintze, A. (2016) Evolutionary game theory using agent-based methods. *Physics of life reviews*, **19**, 1–26.
- [30] Couzin, I. (2007) Collective minds. Nature, 445, 715.
- [31] Li, C., Liu, F. & Li, P. (2018) Ising model of user behavior decision in network rumor propagation. *Discrete Dynamics in Nature and Society*, **2018**.
- [32] Mahmoodi, K. & Gonzalez, C. (2019) Emergence of collective cooperation and neworks from selfishtrust and selfish-connections. Cognitive Science (The 41st Annual Meeting of the Cognitive Science Society), https://mindmodeling.org/cogsci2019/papers/0392/0392.pdf, 2254-2260.
- [33] Gintis, H. (2014) The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences-Revised Edition. *Princeton University Press.*
- [34] Gonzalez, C., Mahmoodi, K. & Singh, K. (2020, in preparation) Emergence of Collective Cooperation in the Absence of Interdependency Information.
- [35] West, B.J., Turalska, M. & Grigolini, P. (2014) Networks of Echoes, Imitation, Innovation and Invisible Leaders. *Springer, New York.*
- [36] Grinstein, G., Jayaprakash, C. & He, Y. (1985) Statistical mechanics of probabilistic cellular automata. *Physical review letters*, 55, 2527.
- [37] West, B.J. (2017) Nature's Patterns and the Fractional Calculus. Walter de Gruyter GmbH & Co KG.
- [38] Bettencourt, L.M. (2013) The origins of scaling in cities. Science, 340, 1438–1441.



FIG. 12: The colors on the top and bottom panel show the average of the ratio of Mutual Cooperation (RMC)at time 100, when each agent played the PDG for 100 times with its eight nearest neighbors, and updated its decision by imitating the decision of the richest neighbor (including themselves). The agents were located on a regular two-dimensional lattice of size 10×10 (top panel) and 30×30 (bottom panel). The horizontal axis of the panels shows the degree of temptation to cheat experienced by the agents and the vertical axis is the ratio to initial cooperators on the lattice. 100 ensembles were was used to evaluate the average RMC at time 100.

Appendix A: Importance of network reciprocity on survival of cooperators in Nowak and May model

To show the importance of the specific connection of agents on a lattice in Nowak and May [24] model, we replicated their simulation work. In this simulation the agents are located on a regular two-dimensional lattice, each having eight nearest neighbors to play the PDG. There is an agent located at each node of the lattice

and they each update their decision at each round of the computation by echoing the decision of the richest agent in their neighborhood (including themselves). Figure 12 shows our replication of their work. The colors on the panels indicate the average of the Ratio of Mutual Cooperation (RMC) which sustained between the agents located on the 10×10 lattice (top panel) and on the 30×30 lattice (bottom panel) after each agent played 100 times with its eight nearest neighbors. The yellow areas correspond to high RMC which happened for low temptation to cheat T_c (i.e., agent's selection of the Defect action in the PDG) and high initial number of cooperators. The yellow area is more extended in the case of agents located on the larger lattice of size 30×30 (bottom panel).

To explicitly show the importance of Nowak and May's assumption of the lattice structure for survival of mutual cooperation, we ran their model and compared time evolution of the RMC between these agents in the case where the agents were paired up randomly. The top panel of Figure 13 shows the ensemble average of 100 simulations for 100 agents (black curve) and 900 agents (red curve) connected on a two-dimensional lattice. At each time round, each agent plays the PDG, with a low temptation to cheat $T_c (= 0.25)$, with all its eight neighbors. Initially 75% of the agents were cooperators, but the RMCevolved and sustained in both cases. The bottom panel of Figure 13 shows the results of the same experiment, but selecting the interacting pairs randomly (no lattice structure is assumed for the agents). As shown in the figure, the RMC vanishes in the absence of the network structure.

Appendix B: Summary of the DDM and SOTC models

The decision making model (DMM) is based on N^2 two-state agents placed at the nodes of a two-dimensional $N \times N$ lattice with periodic boundary conditions, thereby constituting a torus on which the dynamics take place. In the absence of interactions the agents have an exponential probability of switching between the two states (+,-) at a constant rate. If K is the strength of the interaction between agents, $N_i(t)$ is the number of nearest neighbors in the state j = + or -, at time t and $exp[-K[N_+(t) - N_-(t)]/4]$ is the modification of the transition rate for an agent. Note that this timedependent rate can be different for each agent and the probability of switching states at each point in time can increase or decrease depending on whether $N_{+}(t)$ is greater or less than $N_{-}(t)$ for that agent. The DMM is computationally intensive and involves solving an N dimensional two-state master equation. However, because the interactions are local, symmetric and random is sufficient to prove that the DMM belongs to the universality class of kinetic Ising models. Consequently, there is a critical value for K at which the group undergoes a phase transition and reaches consensus. Other interest-



FIG. 13: Top panel: The time evolution of the Ratio of Mutual Cooperation (RMC) for agents located on a regular two-dimensional lattice with 10×10 (black curve) and 30×30 agents (red curve). Initially 75% of agents randomly picked to be cooperators and at each time agents played PDG with $T_c = 0.25$ with their 8 nearest neighbors and imitated the decision of its richest neighbor (including themselves). Bottom panel: The time evolution of the RMC for the agents with similar condition as those adopted in the top panel

except that at each time two agents are picked randomly and played PDG (no predefined connections). The curves are averaged over 100 simulations.

ing properties, including the sensitivity to perturbations, the robustness of configurations to zealots, are recorded and discussed in West et al. [35].

The DMM was generalized to form the self-organized temporal criticality (SOTC) model, which has a form of social interaction that takes place at the level of individuals and is not externally imposed as it was through the choice of K in the DMM. This was accomplished by making the interaction strength time-dependent through



FIG. 14: Flowchart of the SA. "Y" and "N" letters represent "Yes" and "No", respectively.

the introduction of a second network on which the PDG is played. The resulting internal dynamics of the twolevel network in the SOTC model spontaneously seeking consensus [8, 20]. This bottom-up process is not only in the best interest of the individual, but provides optimal benefit for society as well.

Appendix C: Details for the SA algorithm

The following steps are executed during a single cycle initiated at time t of the SA algorithm depicted in Figure 14

a. Step 1. Selfish algorithm - connection (SAC)Agents *i* and *j* are picked randomly such that at time *t* the SAC agent *i* has the propensity:

$$P_{ij}(t) = \frac{M_{ij}(t)}{\sum_k M_{ik}(t)} \tag{2}$$



FIG. 15: The curves show the size dependence of the Ratio of Mutual Cooperation (RMC) for the agents used SAL (Top-left panel), SALT (Top-right panel), SALC (bottom-left panel), and SALTC (bottom-right panel) to update their decisions. $T_c = 0.9$ and $\chi = 200$. The blue, red, yellow, purple, and green colors represent systems with 10, 20, 30, 40, and 50 agents, respectively. The curves are ensemble averages over 10 simulations.

to play with agent j, where the propensity falls in the interval $0 < P_{ij}(t) < 1$. The quantity $M_{ij}(t)$ is the cumulative tendency for agent i to select agent j to play at time t. This cumulative tendency changes at step 7, according to the last two payoffs received by agent i.

At the same time, the SAC agent j has a propensity given by Eq. (2), with the indices interchanged, to play with agent i, where the propensity falls in the interval $0 < P_{ji}(t) < 1$. Two agents i and j partner if two numbers r_1 and r_2 are randomly chosen from the interval (0,1), and satisfy inequalities $r_1 < P_{ij}(t)$ and $r_2 < P_{ji}(t)$. If both inequalities are not satisfied another two agents are randomly selected at each time t until the inequalities are satisfied and the two agents "agree" to play a PDG. In the flowchart wherever letter r is called it returns a single uniformly distributed random number in the interval (0,1). Notice that if step 7 is inactive, then step 1 turns to simply picking two random, different agents.

b. Step 2. Selfish algorithm - learning (SAL) Agent i, playing with agent j, initially selects an action, C or

D, using SAL. The agent i has the propensity:

$$PC_{ij}(t) = \frac{C_{ij}(t)}{C_{ij}(t) + D_{ij}(t)}$$
(3)

to pick C and the propensity:

$$PD_{ij}(t) = \frac{D_{ij}(t)}{C_{ij}(t) + D_{ij}(t)}$$
(4)

to pick D, as it's next potential decision. Note that the sum of these two terms is one. The quantities $C_{ij}(t)$ and $D_{ij}(t)$ are cumulative tendencies for agent *i* playing with agent *j*, at time *t*, for the choice *C* or *D*, respectively. These cumulative tendencies change at step 6 based on the last two payoffs of agent *i*.

To decide on an action a random number r is selected from the interval (0,1). If $r < PC_{ij}(t)$ then the next decision of SAL agent i will be C, otherwise will be D. The same reasoning applies for agent j. c. Step 3. Selfish algorithm - trust (SAT) Instead of executing the decision determined by SAL in step 2, agent *i* has a propensity to trust the decision made by agent *j*, which also used SAL in step 2. The propensity for agent *i* to trust the decision of agent *j* is:

$$PT_{ij}(t) = \frac{T_{ij}(t)}{T_{ij}(t) + \overline{T}_{ij}(t)}.$$
(5)

Here again, if a random number r, chosen from the interval (0,1), is less than $PT_{ij}(t)$ then trust is established. The quantities $T_{ij}(t)$ and $\overline{T}_{ij}(t)$ are cumulative tendencies for agent i to execute the choice of agent j, at time t, or to not rely on trust and to execute its choice based on SAL, respectively. These cumulative tendencies update in step 5 based on the last two payoffs of agent i.

d. Step 4. Evaluating Own Payoffs After agent iand j executed their action, C or D, their payoff is evaluated using the payoffs matrix of the PDG, $\Pi_{ij}(t)$ and $\Pi_{ji}(t)$, respectively.

e. Step 5. Update of cumulative tendency of SAT If agent i used SAT playing with agent j then the accumulative tendencies T_{ij} and \overline{T}_{ij} change to $T_{ij} + \Delta_{i,jk}$ and $\overline{T}_{ij} - \Delta_{i,jk}$ for the next time agent i and j partner. The same happens for agent j. Similarly, if agent i did not use SAT, but relied on its own SAL, then the accumulative tendencies \overline{T}_{ij} and T_{ij} change to $\overline{T}_{ij} + \Delta_{i,jk}$ and $T_{ij} - \Delta_{i,jk}$ for the next time agent i and j partner. The same happens for agent j.

f. Step 6. Update of cumulative tendency of SAL Step 6 is only active if the agent did not use SAT at step 3. If agent *i* played *C* with agent *j*, then the accumulative tendencies C_{ij} and D_{ij} change to $C_{ij} + \Delta_{i,jk}$ and $D_{ij} + \Delta_{i,jk}$ for the next time agent *i* and *j* partner. If agent *i* played *D* with agent *j*, then the accumulative tendencies D_{ij} and C_{ij} change to $D_{ij} + \Delta_{i,jk}$ and $C_{ij} - \Delta_{i,jk}$ for the next time agent *i* and *j* partner. The same happens for agent *j*.

g. Step 7. Update of cumulative tendency of SAC In this step the cumulative tendency to play with a specific agent changes. If agent *i* played with agent *j* then the cumulative tendency of pairing with agent *j*, M_{ij} changes to $M_{ij} + \Delta_{i,jk}$. The same happens for agent *j*.

As boundary condition for steps 5, 6 and 7, if the updated cumulative tendency goes beyond its defined maximum (or below its defined minimum) then it has to set back to the maximum value (minimum value).

h. Boundary condition We set the maximum and minimum of the cumulative tendencies to be 1000 and 0, respectively. If an updated cumulative tendency went above 1000 (or below 0) then it set back to 1000 (or 0).

Appendix D: Robustness and Sensitivity Tests

We ran a number of simulations to test the robustness of the SA algorithm. First, Figure 15 shows the effect of the size of the system (M = 10, 20, 30, 40 and 50) on the *RMC* for different mechanisms of the *SA*. As observed, regardless of the system size the *RMC* curves converge; although it takes longer the more agents there are in the system. Thus, although size of the system alters the learning time, it does not affect the main result of agents' convergence to cooperation.

Second, we tested the effect of the payoffs of the PD matrix on the emergence of cooperation. We set $T = 1 + T_c = 1.9$, R = 1 and S = 0 but changed the values of P (in the interval of [0, 0.2]). For each P we ran ten simulations and evaluated the RMC at $t = 10^6$ for SAL, SALT, SALC, and SALTC mechanisms. The top panel of Figure 16 shows that in the case of SAL, where there is just one learning mechanism at work, the level of RMC decreased steadily as the payoff for DD (P) increased. This is due to formation of DD reciprocity between the agents; Increasing P alters the set of $\Delta_{i,ik}$ s, that agents use to updated their cumulative propensities, in favor of increase of mutual defection DD. For SALCmechanisms, the decrease in RMC is more as the connection mechanism improved the DD reciprocity, which could dominate the CC reciprocity. In other words, playing most of the time with one player prohibited the agents from exploration and finding the advantage of CC over DD. Both SALT and SALTC (all the mechanisms of the SA active) could sustain the level of RMC as P increased. This shows the advantage of learning through trust in the SA. Trusting the decision of other agents improved the exploration capability of the agents to learn the advantage of CC over DD.

Third, we tested the sensitivity of our results to the value of χ . The bottom panel of Figure 16 shows the results of the simulation with a higher value of $\chi = 400$. As observed, increasing χ improved the level of RMC for all the mechanisms. Higher coefficient χ increased the magnitude of the $\Delta_{i,jk}$ s; increased the exploration capability of the agents.

Appendix E: Effect of trust (fixed chance of trust) on the evolution of decisions

The aim of this section is to highlight the difference between trust used in the literature with our dynamic trust. We introduced "trust" and "not trust" as a "decision" (while the word trust means "not making a decision"). We connected trust to a reinforcement learning mechanism to make it adaptable. To show the difference, here we determine the effect of fixed chance of trust between the agents on the dynamics of decisions where there is no reinforcement learning mechanism through SAT. We investigated the dynamics of the decisions made by social groups consisting of 10, 20, or 30 agents. These agents randomly partner at time t and can either retain their decision, or exchange it with a fixed chance for their part-



FIG. 16: The top panel shows the effect of changing the value of P (the value of punishment in the PD game) on the Ratio of Mutual Cooperation (RMC) at time 10^6 for different mechanisms of the SA. For each value of P and each mechanism there are ten points resulted from ten simulations. The colors red, black, blue and green correspond to the SALTC, SALT, SAL, and the SALC mechanisms, respectively. The bottom panel shows the effect of increasing the value of χ to 400 on the results of the top panel. $T_c = 0.9$

ner's decision (trusting). In top panel of Figure 17 the average time for the agents to reach consensus is plotted versus the chance of trust that agents have to take the decision of their partner as their next decision in each of the three groups. This figure shows that there is a broad flat minimum (optimum) value of trust that leads the agents to achieving consensus faster and that this time increases with the size of the network.

It is interesting to note that the evolution of mutual cooperation when one of the agents becomes a zealot, which is to say, that agent remains the same all the time, the network rapidly relaxes to the value of the zealot. The conclusion is that this network is highly sensitive to this small perturbation depending on the chance of trust.



FIG. 17: Top panel: The black, red and blue curves show the average time to reach consensus (the time that all the agents reach in +1 or all in -1 state) vs. the magnitude of the trust between them for systems with 10, 20, and 30 agents, respectively. At each time an agent can keep its previous decision or can imitate the decision of its partner by the fixed propensity of trust. Bottom panel: Dependence of the average time to consensus to the number of the agents with 0.1 (black), 0.5 (red) and 0.9 chance of trusting the decision of their pair. The power law coefficients for N^{α} are 2.37, 2.28 and 2.29, respectively. The curves are ensemble averages over 100 simulations.

In other words, this system, that learns only through trust is not significantly resilient (is not robust).

The panel on the bottom of the Figure 17 indicates a dependence of the average time for a group to reach consensus from a random initial state is a monotonously increasing function of group size N. The average time to reach consensus satisfies an allometry equation [37]: $Y = aX^b$, where Y is the functionality of the system, X is a measure of system size and the empirical parameters a and b are determined by data. Allometry relations (ARs) have only recently been applied to social phenomena [38] but on those applications the scaling has always been superlinear (b > 1). Bettencourt [38] point out that the superlinear scale reflects unique social characteristics with no equivalent in biology in which the ARs scaling is invariably sublinear (b < 1). They point out that the knowledge spillover in urban settings drive the growth of functionality such as wages, income, gross domestic product, bank deposits, as well as rates of invention...all scale superlinearly with city size. In this paper the agents were allowed to learn through a variety of modalities and to find their individual optimum value of trust regarding other agents. Subsequently, we showed that this dynamic trust acting, in concert with other learning modalities, leads to robust mutual cooperation.