



OPEN

Dynamics of collective action to conserve a large common-pool resource

David Andersson^{1,2}, Sigrid Bratsberg², Andrew K. Ringsmuth^{3,4,5,6} & Astrid S. de Wijn^{1,2}✉

A pressing challenge for coming decades is sustainable and just management of large-scale common-pool resources including the atmosphere, biodiversity and public services. This poses a difficult collective action problem because such resources may not show signs that usage restraint is needed until tragedy is almost inevitable. To solve this problem, a sufficient level of cooperation with a pro-conservation behavioural norm must be achieved, within the prevailing sociopolitical environment, in time for the action taken to be effective. Here we investigate the transient dynamics of behavioural change in an agent-based model on structured networks that are also exposed to a global external influence. We find that polarisation emerges naturally, even without bounded confidence, but that for rationally motivated agents, it is temporary. The speed of convergence to a final consensus is controlled by the rate at which the polarised clusters are dissolved. This depends strongly on the combination of external influences and the network topology. Both high connectivity and a favourable environment are needed to rapidly obtain final consensus.

Common-pool resources (CPRs) are ‘non-excludable’ (open-access) and ‘subtractable’ (one individual’s consumption can preclude another’s). Harvesting a CPR for short-term benefit without restraint can make it unavailable through depletion—a so-called ‘tragedy of the commons’—unless harvesters can organise to restrain their harvesting to a sustainable level.

CPR management has typically been studied in systems where harvesters can perceive the resource level’s linear response to their harvesting, allowing them to adapt efforts accordingly (e.g. pastures, groundwater, fisheries). In models, this appears as explicit interdependence between harvesting strategy evolution and resource level^{1–6}. However, a CPR’s linear response may be imperceptible to harvesters if, for example, the resource pool is very large when compared with the harvesting rate. Such a resource may show no obvious reason for restraint, even after extensive harvesting. However, the state of the encompassing social-ecological system may depend nonlinearly on the resource level, so unrestrained harvesting may push the system to an unexpected critical transition into a new and potentially undesirable state⁷. Three prominent examples of such CPRs are unpolluted atmosphere, global biodiversity and public healthcare systems during a pandemic.

Conservation of a large CPR can be considered as a large-scale (potentially global) collective action dilemma^{8,9}. In such a dilemma, group members must choose whether to contribute to a collective action such as adherence to a pro-conservation behavioural norm which, if successful, will benefit the whole group. Below some threshold rate of cooperation, each individual benefits more from defecting than cooperating, but all individuals are better off if the threshold is reached¹⁰.

The lack of obvious feedback from a large CPR amplifies the collective action problem because the need for restraint may not be obvious to enough people until tragedy is (almost) inevitable. Solving this problem requires achieving a sufficient level of cooperation with a pro-conservation behavioural norm fast enough for the collective action to successfully conserve the resource. Precisely what this level and rate are depends on the details of a given CPR and how it responds to changes in harvesting behaviour. In all cases, however, the behavioural change process will depend on agents being affected by information received from other agents and from the broader sociopolitical environment. In general, the faster and more complete this process is, the more likely it is that the conservation action will be successful.

¹Chemical Physics Division, Department of Physics, Stockholm University, Stockholm, Sweden. ²Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway. ³Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden. ⁴Section for the Science of Complex Systems, Medical University of Vienna, Vienna, Austria. ⁵Present address: Complexity Science Hub Vienna, Vienna, Austria. ⁶Wegener Center for Climate and Global Change, University of Graz, Graz, Austria. ✉email: astrid.dewijn@ntnu.no

Models of opinion dynamics on social networks have been extensively studied^{11,12}. Many are agent-based and assume ‘bounded confidence’: an agent’s interaction with another may change its opinion but only if they engage meaningfully. Due to the psychosocial phenomenon of homophily, this is unlikely to happen between people who have very different views. Bounded confidence models^{13,14} therefore assume that agents engage only if their opinions differ by less than some specified threshold. In contrast to this psychology-based approach, models of CPR management typically assume that agents respond rationally to available information in a manner described by evolutionary game theory. The two types of model can behave quite differently.

Large-scale CPR conservation efforts face psychosocial constraints and are also rationally motivated. They thus combine features of the above two model types. Agents will face material consequences in the long term and at least some people are aware that it is in their long-term self interest to try to mitigate these consequences through some short-term sacrifices. They might also realise that the collective action problem may be solved more quickly if they are willing to discuss with others. This means actively overcoming bounded confidence to engage with people around them even if they have very different views.

While long-term limiting (equilibrium) behaviour in both CPR problems and opinion dynamics has been widely studied^{15,16}, the transient dynamics have received far less attention. Successful conservation, however, may depend not only on the long-time group behaviour but also on the rate of behavioural change once the problem is first recognised. A small number of opinion dynamics studies have specifically modelled transient dynamics^{17,18} and provided evidence that this is important for understanding realistic systems of practical interest, which are often in such transient states.

In this work, we study the transient dynamics of collective action through behavioural change mediated by social interactions between agents who are also influenced by a broader sociopolitical environment. Being rationally motivated by the need for action, neighbours of different opinions overcome bounded confidence and are just as likely to interact with each other as neighbours of similar opinions. We also expose the agents *en masse* to a global field that represents the combined effect of influences in the broader sociopolitical environment. We study the competition between these local and global influences, and how the speed of consensus formation emerges from different features of the model. Finally, we discuss implications of our findings for real-world management of large CPRs.

We study the spread of cooperation with a pro-conservation behavioural norm using a model that is similar to those used for opinion dynamics without bounded confidence¹⁹. Each agent’s level of cooperation is represented by a real number from -1 to 1 , which respectively represent total defection and total cooperation, and we refer to agents at these extremes as defectors and cooperators accordingly. We adopt this nomenclature from evolutionary game theory even though our model more closely resembles an opinion dynamics model, to reflect the mixture of rationally motivated, strategic behaviour and psychosocial constraints in the norm-spreading dynamics that we aim to capture. Since we are concerned with building collective action from a state of inaction, our simulations begin with network states in which most agents defect.

We study dynamics on clustered scale-free networks. Real social acquaintance networks are typically clustered scale-free (CSF) networks^{20,21}. In such networks, topological path lengths between agents are typically short (even in the case of large and highly connected online social networks²²), giving rise to the name ‘small-world network’²³. Although scale-free networks have been extensively studied²⁴, frameworks for constructing and analysing clustered scale-free networks were developed only recently. In addition to their small-world properties, such networks have high connectivity between proximate nodes (i.e. topological clustering). However, we caution the reader that, except where referring to CSF networks directly, in our analysis we use ‘cluster’ not in this topological sense but rather to refer to a group of agents who share a common cooperativity (i.e. strongly correlated states). We study collective dynamics across many simultaneously evolving network realizations and average the resulting network properties over the many realizations.

Results

Figure 1a shows a typical network evolution, averaged over 500 network realizations. Parameters are standard (see “Methods” section as well as Supplementary Material) except $\phi = 0.015$ for the external field. In this and similar plots, we present the average cooperativity and standard deviation. The various parameters are defined in the “Methods” section and summarised in Table 1. The standard deviation here is calculated by first calculating the standard deviation of the cooperativity in each individual network realization, and then averaging these.

Transient polarisation. The dynamics show two characteristic time scales. Initially, we see a rapid decline in the average cooperativity and increase in its standard deviation. However, the dynamics soon become dominated by a second, slower process that gradually drives the network to a fully cooperative state. We can understand this better from simulations on smaller 13×13 grids with the same degree and periodic boundaries, shown in Fig. 1, along with the evolution of the average cooperativity and standard deviation. Initially, neighbouring agents are uncorrelated. Agents that are initially strongly cooperative or defective spread their behaviours to their surroundings and strengthen their own behaviours. Correlated clusters emerge around the original agent and any interaction within the cluster will serve only to reinforce the behaviours, leading locally to an ‘echo chamber’ and globally to a strongly clustered, polarised network. Once the network has become completely polarised, meaningful interactions take place only on the boundaries separating opposing clusters.

Comparing the snapshots in Fig. 1 to the time evolution shows that this cluster formation regime corresponds to the first stage in Fig. 1a. Since the network begins with a small majority of defectors, they on average spread their behaviours more effectively than cooperators. Hence we see strong initial decrease in the average cooperativity, averaged over many networks. Notably, such a polarised state is often the equilibrium to which bounded-confidence models with a sufficiently restrictive threshold converge. In our system, however, it is only

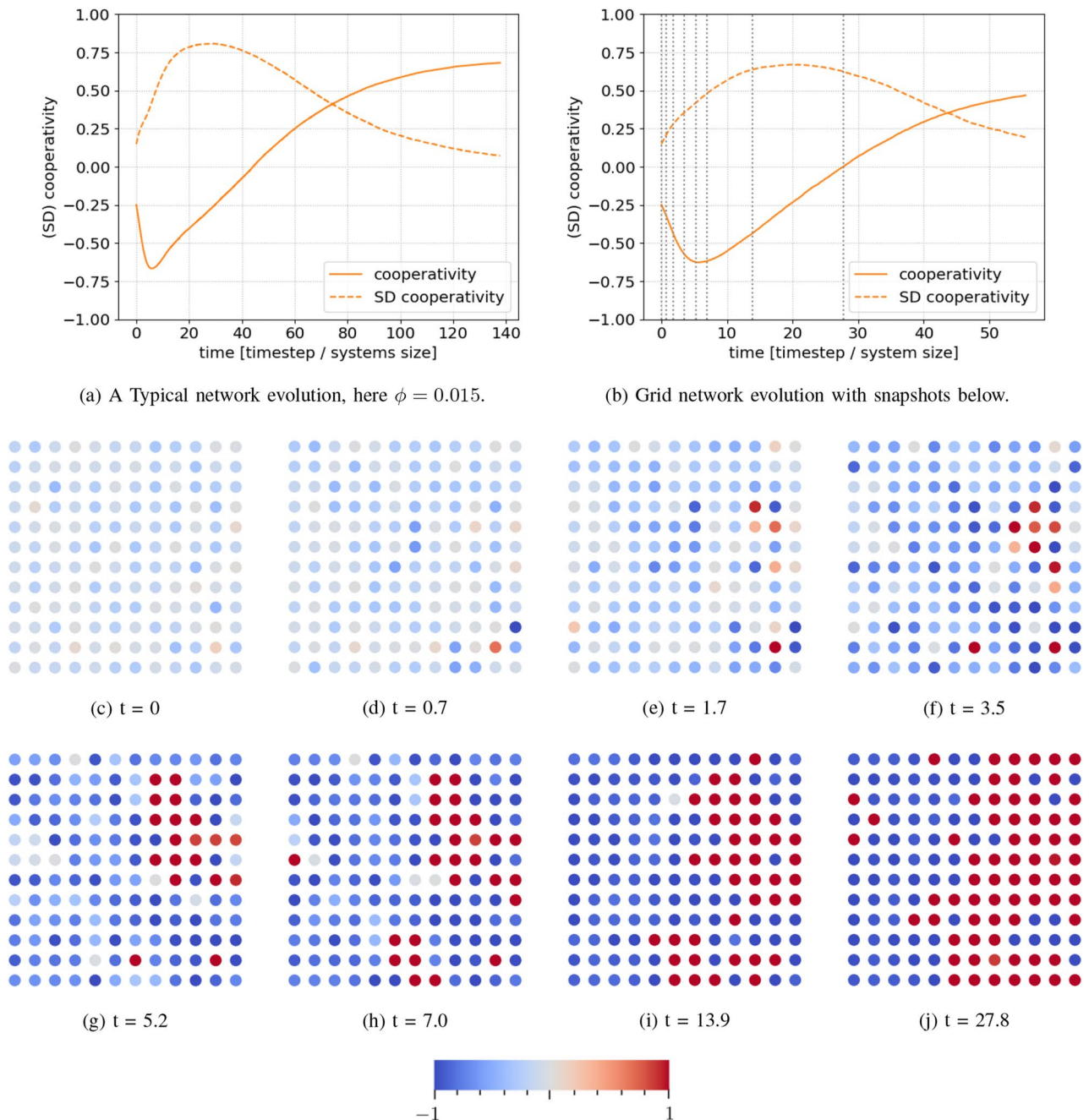


Figure 1. (a) A typical average time evolution of the cooperativity for conditions slightly favouring cooperation on a CSF network. At first, the short range but strong social pressure of the initially abundant defectors dominate the evolution. However, after some time, the long-range (global) but weak external field drives the networks slowly into cooperation. The standard deviation initially increases during cluster formation, but later decays away as the network homogenizes. (b) An example of how a grid network of small size evolves. Comparing (b) with grid snapshots (c–j) helps to understand how the clustering process evolves; the dashed lines in (b) indicate the times corresponding to (c–j). Moreover, comparing (b) with (a) shows the similarity of dynamics in the grid and CSF network cases. (c–j) A realization of a small grid network with periodic boundary conditions and an initial defector majority, but a weak field favouring cooperation. Blue indicates defective behaviour, red cooperative. We see that, initially, agents with strong opinions efficiently convert undecided agents in their vicinity to their cause. This is the cluster-formation regime of the dynamics. In this realization, one can observe a strongly cooperating (red) agent in the upper right, which appears at $t = 1.7$. This agent then forms a large cooperative cluster around it. In contrast, most of the rest of the network gradually becomes more defective due to the initial abundance of defectors. At $t = 13.9$, the network is almost completely polarised, and by $t = 27.8$, the cooperative cluster has begun to overcome the defective cluster by gradually dissolving its perimeter.

Parameter/variable	Represents	Value
Agent behaviour (b)	The strategy (cooperative or defecting) that an agent operates	–
External field (ϕ)	Global factors	0.05
	Policy regulation, mass media, morals, etc	
Self weight (w_i)	Resistance to influence	0.6
	Stubbornness, ideological commitment, etc	
Agent-agent weight (w_{ij})	Susceptibility	0.5
	Friendship, animosity, charisma, etc	
Agent-agent weight STD	Susceptibility variance	0.15
Initial cooperation ratio	Initial average cooperativity	– 0.25
Initial cooperation STD	Variance in initial cooperativity	0.15
Randomness (r)	Noise intensity in agent interaction	0.1
Growth degree	Degree of nodes added during network growth as introduced by the Holme–Kim algorithm	8
Clustering	Relative level of clustering of networks as introduced by the Holme–Kim algorithm	0.5

Table 1. Summarizing all model parameters and variables to provide an overview of the model's components. Here the reference value of each parameter is given, which is used throughout this paper unless otherwise explicitly stated.

transient. Once polarisation is complete and no or few neutral agents remain, a ‘tug-of-war’ takes place between clusters, at their edges. In this competition, a sufficiently strong pro-conservation external field can tip the balance, making it more likely that one-on-one interactions will result in defectors converting to cooperators, rather than the converse. We see this in the second dynamical stage, in which the cooperativity slowly rises as defective clusters shrink.

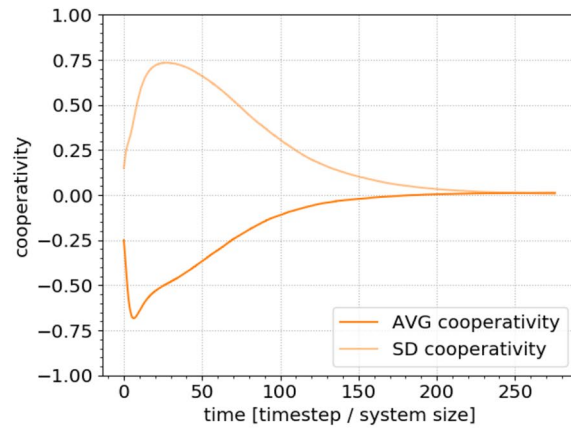
Figure 1a shows that the point of maximum polarisation, where the standard deviation is maximal, does not coincide exactly with the point of minimal cooperation. In Fig. 1c–j, we see that, at the cooperation minimum, there are regions that are not yet polarised but are responding to the pro-conservation external field. The dynamical crossover from rapid polarisation to slow cluster growth/shrinkage thus happens somewhat after the minimum. A consequence of this is that the average cooperativity is not bound to monotonically increase after the apparently minimal average cooperativity is reached. Rather, at the end of the cluster formation stage, in some cases, even though the average cooperativity has already started to increase slightly, the cooperators nevertheless are not abundant enough and the system converges to a fully defective state (Fig. 2b).

We can understand the above results in more detail by considering the strengths of the two competing interactions. More-or-less neutral agents are more susceptible than polarised agents to persuasion by others. Since neutral agents are more abundant in the initial network configuration, social interactions more strongly affect its early dynamics, which quickly converge to a polarised state. Once the system is strongly polarised, persuasion is possible only between agents of extreme, opposite levels of cooperativity at the cluster boundaries. In this situation, even a small contribution from the external field can tip the probability in one direction over the other. Thus, the defector is slightly more likely to become a cooperator than the cooperator is to become a defector. When this is the dominant dynamics, the network evolves more slowly, but the influence of the external field becomes apparent.

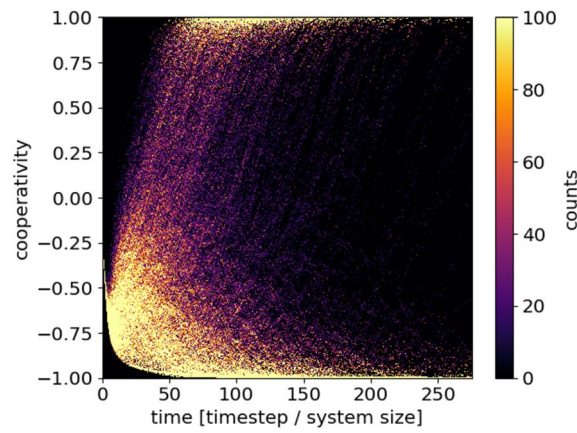
In some regions of the parameter space, we observe so-called social tipping points, around which the network's long-time behaviour depends sensitively on the parameters. Figure 2 shows a network evolution around such a tipping point. In panel (a), both the average and standard deviation of the cooperativity converge to roughly zero, implying that the individual network realizations converged to fully cooperative and fully defective states with roughly equal frequencies. Panel (b) shows a 2D histogram, binning the cooperativities of the individual network realizations that are averaged in (a). The colour scale is capped at 100 counts to enhance contrast. We see a clear division of the network realizations between the two attracting terminal states. The final outcome for each realization depends on its precise details. The initial conditions were chosen such that fully cooperative and fully defective outcomes are both likely. The external field here was $\phi = 0.0103$, and all other parameters standard.

Convergence rates. Figure 3 gives more insight into the process of convergence to a fully cooperative network by plotting the same transient dynamics as Fig. 1a on a logarithmic scale. We see that the second, convergent dynamical stage in Fig. 1a is itself separated into two stages. The first is an approximately exponential stage of convergence, as expected. In the second, slower, stage, it is likely the finite size of the remaining clusters that plays a role, in combination with fluctuations.

In real-world collective action problems, achieving the threshold needed for effective action is far more important than converting the last few hold-outs. Therefore it is the exponential stage of the convergence regime that interests us most. Since the length of this stage varies, we choose a simple and consistent way to evaluate the rate: taking the time derivative of the average cooperativity when the average cooperativity is close to 0, where the curve is certain to be exponential.



(a)



(b)

Figure 2. Long-term evolution of a system close to a tipping point ($\phi = 0.0103$). (a) The average and standard deviations of the cooperativity converge to a value near zero, meaning that in roughly half of the realizations, the networks became completely cooperative, and in the other half, completely defective. This evolution is very sensitive to changes in parameter values around this point, as is a hallmark of tipping points. (b) A heatmap corresponding to all the data for which the averages are shown in (a). The intensity corresponds to the density of networks having a specific average cooperativity at a specific time. Most networks converge to full cooperation or defection within the simulation time, reinforcing that this is a tipping point.

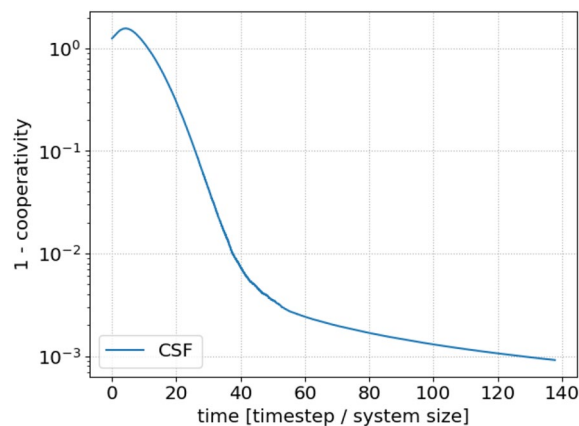


Figure 3. Logarithmic plot for a long range simulation of the average cooperativity in a standard clustered scale free network. We see that there is a critical slowing down at extremely homogeneous systems.

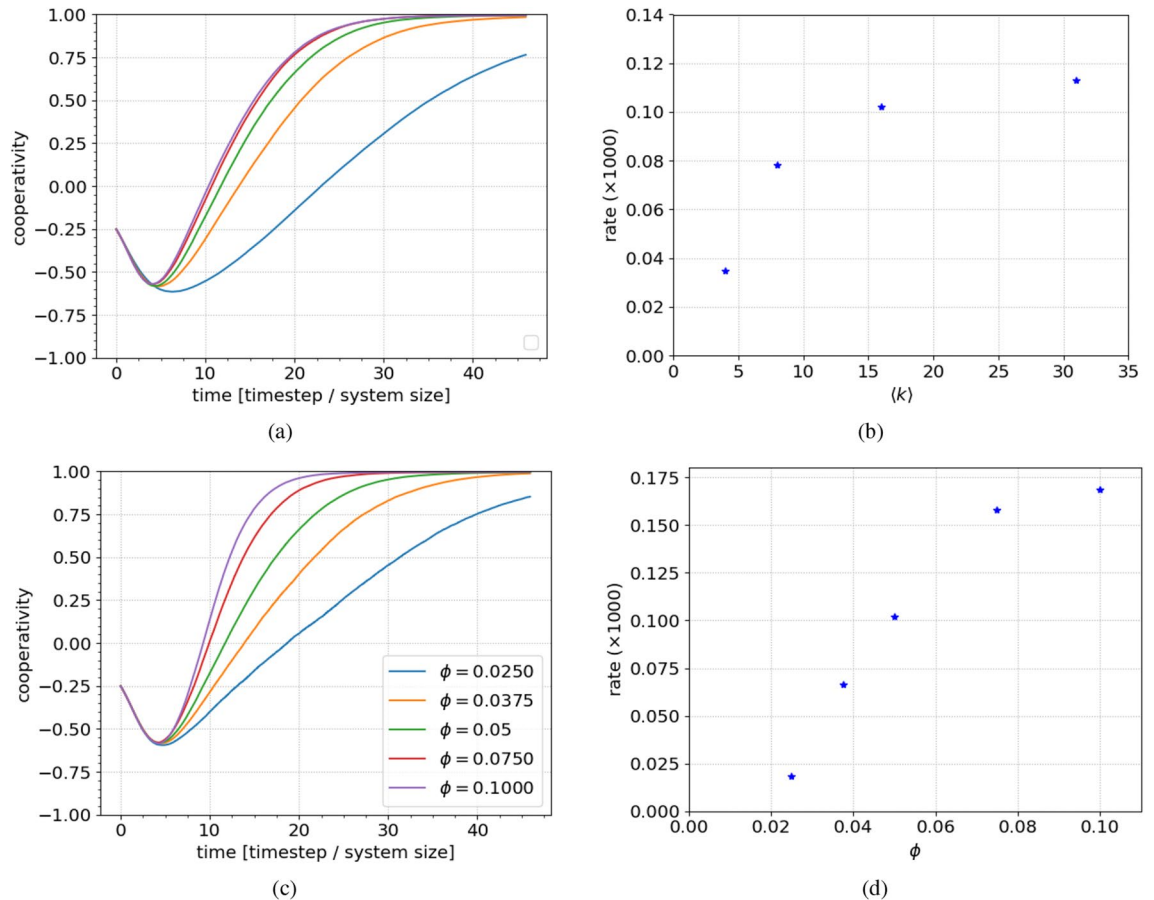


Figure 4. The effect of the connectivity and field on the convergence to the final state. **(a,b)** The average cooperativity as a function of time for different network connectivities and corresponding rates. **(c,d)** The same for field strength. Increasing the average connectivity in the network increases the convergence speeds, because each individual agent on average will have more ways of receiving the dominant message. As expected from the model formulation and the analysis thereof, the strength of the external field is paramount to both the convergence speed as well as the fate of the networks. We see that initially, improving the conditions of the networks yields a significant change in convergence rate. Whereas for already favourable conditions, there is a diminishing return on the gain in convergence rate.

Network connectivity. We investigate how the transient dynamics depend on network topology by comparing the convergence for different networks. Figure 4a shows the average cooperativity evolution for clustered scale-free networks with different degrees and otherwise identical, standard parameters. We see that network connectivity is a crucial determinant of the time scale at which clusters are dissolved. In Fig. 4b, we show the convergence rate as a function of the average degree (obtained by changing the growth degree, k_0 , to 2, 4, 8, 16, 32 respectively, see “Methods” section). Once the network is sufficiently connected, the convergence rate saturates. This is most likely due to the finite size of the network. We can calculate the average number of next neighbours as the square of the average degree. Since our system contains around 1000 agents, if the average degree is 31 or 61, every agent is likely to be connected to a quarter of, or all (respectively), other agents, through its nearest neighbours.

External field. We explore how the external field affects the convergence rate by first letting the network polarise (reach its minimum average cooperativity), and then changing the field. In this way we can discount any effects of the field on the initial polarisation stage as significant determinants of the convergence dynamics.

Figure 4c shows the results. Apart from the delayed change to the external field, the parameters and simulation protocols were identical to earlier simulations. Each curve corresponds to 500 network realizations. As shown in Fig. 4d, the external field strength strongly affects the average cooperativity’s convergence rate, though the response diminishes as field strength increases.

Influencers. Over recent years, it has become clear that so-called *influencers* can affect the spread of behaviours through social networks, particularly on social media. We represent such an influencer in our model as an agent who has both exceptionally high connectivity and strong, inflexible behaviour. Again, we first let the networks polarise and then begin our analysis by identifying the agent with the highest connectivity in the network and

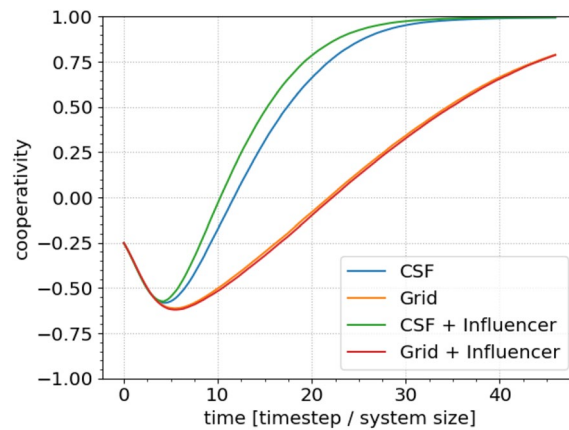


Figure 5. The average cooperativity as a function of time for different grid types with and without an *influencer* in the network. An influencer is a highly connected (in this case the most connected) node with an unwavering behaviour. We can see that the key property of an influencer is the first of these attributes, as their effect in a grid network is almost negligible.

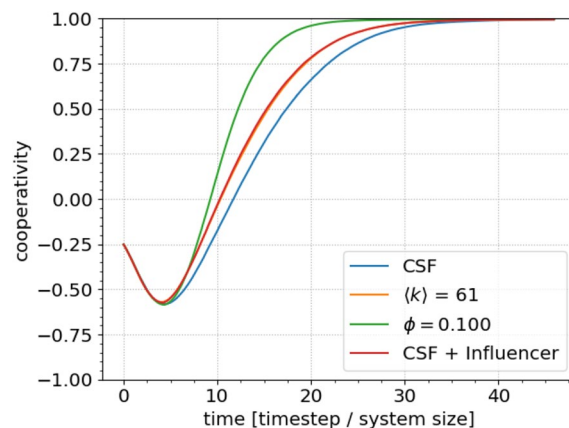


Figure 6. Summarizing cooperativity as a function of time for the various interventions suggested in this paper to increase the rate of convergence towards consensus. Data collected from previously presented figures. The yellow and red curves are overlaid because their impacts are almost identical.

fixing this agent as an unyielding cooperator. To verify that the subsequent dynamics reflect both this agent's behaviour and connectivity, we compare this with the addition of an influencer to a grid network in which all agents are equally well connected. The results are shown in Fig. 5, again with standard parameters and 500 network realizations. Clearly, the presence of an influencer drives substantially faster convergence to the final state.

Figure 6 compares the effects of our various interventions on the convergence rate. For brevity, we include only the strongest case observed for each intervention. Under the given assumptions, the external field most strongly affects the convergence rate. However, we emphasise that this result should be interpreted with caution, since we do not account for factors affecting the practical feasibility of each intervention in any real system. We also leave detailed exploration of interactions between interventions for future work.

Discussion

In this study, we used an agent-based model to investigate the transient dynamics of collective action to conserve a large common-pool resource. This problem combines rationally motivated behaviour with the many unconscious processes intrinsic to social psychology. We used an opinion dynamics-type model, without bounded confidence, to capture this dual nature and studied the effects of network topology, an applied external field, and the presence of 'influencers', on the rate at which the network population converged to a state of full cooperation with a pro-conservation behavioural norm.

Different processes dominate the model on different time scales. Initially, formation of polarised clusters dominates but this gives way to a slower process of cluster dissolution due to social interactions at their perimeters, which depends strongly on cluster geometry. In larger networks, there is a possibility of larger clusters. Since the number of agents inside clusters will grow faster with the cluster size than the number of agents at their perimeters, the convergence will likely be slower.

We found that the convergence rate is strongly affected by both network connectivity and external field strength, and that some parameter regions show tipping points in which the long-time behaviour depends sensitively on the parameters. A pro-conservation field gives cooperators the edge to slowly convert defectors at the cluster perimeters. We also found that a highly-connected and zealous influencer, who can convince others but is itself inflexible, can greatly speed up the convergence.

In the presence of any of these interventions, achieving collective action is essentially a process of dissolving defective clusters, which is fundamentally limited by network topology at the cluster perimeters. This result strongly agrees with recent work by Fotouhi et al.²⁵, who found that deliberately and carefully increasing connectivity between polarised clusters is a powerful mechanism for promoting cooperation. In our view, this speaks particularly to the power of now-ubiquitous social media echo chambers as inhibitors of adequate collective action on the extremely urgent global challenges of climate change and biodiversity loss. Our results suggest that replacing the mechanisms that generate and maintain such polarisation with others that instead promote constructive engagement between disagreeing groups could be a powerful promoter of large-scale collective action. This is an important subject for further modelling, including additional features of real social networks such as link rewiring, which can have a significant impact on formation and therefore also dissolution mechanisms for polarised social clusters^{2,26}.

Previous investigations have extensively and separately discussed social tipping points⁷, clustering^{13,14}, the role of connectivity²⁵, as well as the impact of external parameters²⁷, all of which are represented in our model. Our results suggest that real-world efforts to promote collective action on large-scale problems can benefit from both improving network connectivity *and* providing a sociopolitical climate which favours cooperation. In this sense, we combine previous results into an overarching framework. We hope that this novel perspective will open new avenues into investigating the social dynamics that are critically important for addressing twenty-first century global challenges.

Our results also suggest that convergence rate, rather than just long-time dynamics, are crucially important for real-world collective action. Much effort has been invested in trying to understand which parameters ensure a favorable final state of a social network. However, for urgent large scale common-pool resource problems, simply concluding that there exists a stable favorable solution that will be reached *eventually* is in practice not sufficient to ensure a favourable outcome, since we also have to ensure that we will arrive there before it is too late. Otherwise the commons may have already sustained significant, or even catastrophic and irreversible, damage. We would therefore argue that the complexity of large-scale common-pool resource management is such that considering a single intervention is insufficient for advising mitigation strategies. Ensuring favourable network connectivity would be moot if a favourable sociopolitical climate is lacking, and vice versa.

Finally a word on the scope of this research. Our model differs from the well-established bounded-confidence models that are conventionally used to model opinion dynamics, and which typically produce final states with clustering. This is founded on the fact that widespread consensus is generally not reached on the vast majority of questions about which people have opinions, such as which TV-show is best or which political party to vote for. However, there are several large-scale societal issues on which opinion and/or behavioural homogenization is occurring, such as gender equality, democratization, and possibly the need for substantial climate change mitigation. All of these issues are characterized by the threat of long-term system-wide material consequences for inaction, giving agents a rational incentive towards cooperation. An interesting avenue for future research would be to introduce a parameter for the level of confidence bounding to our model and investigate the transition between the clustered final state typically seen in regular bounded confidence models and the final full consensus observed in the case of unbounded confidence analysed here. To put things into perspective, it should be noted that bounded confidence models with a large confidence interval can also produce a consensus end state. However, tuning which state becomes the end state is generally not a natural component of such models.

Methods

We study the spread of cooperation with a pro-conservation behavioural norm using a formalism that is similar to other models of opinion dynamics without bounded confidence¹¹. Each agent's level of cooperation (b_i) is represented by a real number from -1 to 1 , which respectively represent total defection and total cooperation, $b_i \in [-1, 1]$. Since we are concerned with building collective action from a state of inaction, our simulations begin with network states in which most agents defect.

The network evolution is simulated iteratively in 4 steps: 1. Randomly select an agent (i); 2. Randomly select a neighbour (j) of i ; 3. Evaluate their interaction as described below. 4. Update i 's behaviour according to the outcome. This process is iterated until the network converges to a steady state. Updating i 's behaviour, rather than j 's, ensures that agents with more connections are more influential.

The change Δb_i in i 's behaviour resulting from an interaction is given by

$$\Delta b_i = |b_i - b_j| [w^+(1 - b_i) - w^-(1 + b_i)], \quad (1)$$

where w^+ and w^- describe the weights of opinion changes in the positive and negative directions respectively. Here $w^- = 1 - w^+$ and $w^+ = f(x_{ij})$ with

$$f(x) = \begin{cases} 0 & \text{if } x \leq -r \\ \frac{1}{2r}(r + \xi) & \text{if } -r < x < r \\ 1 & \text{if } x \geq r, \end{cases} \quad (2)$$

where ξ is a random noise term and furthermore

$$x_{ij} = w_i b_i + w_{ij} b_j + \phi. \quad (3)$$

We can simplify equation 1 to

$$\Delta b_i = |b_i - b_j| [2f(x_{ij}) - b_i - 1]. \quad (4)$$

These equations are motivated as follows. The first is a continuous generalization of a discrete opinion exchange, as reviewed in¹⁹. In such models, one would weigh the probability of an agent flipping its state when evaluating the agent interaction, this is the function of w in Eq. (1). Additionally, in this interaction we have added a prefactor that limits the degree to which two already agreeing agents can influence each other and, conversely, ensuring that an agent has to be presented with a new opinion in order to change its opinion.

Equation (3) describes the magnitude of the agent-to-agent interaction. The function $f(x)$ captures the response of the affected agent, including the random environmental influences, which are in r . Finally, Eq. (4) is just a more compact version of Eq. (1) that is more useful for computations.

We now define every parameter in the model (summarised in Table 1). The parameter, ϕ represents the influence of the sociopolitical environment, which we assume to affect all agents equally. Media influence has been variously studied in opinion dynamics models (e.g.^{27,28}), usually implemented to represent a specific mechanism or scenario. Since our focus is more general, ϕ is intended as a simple representation of the net influence of many factors including media messaging and regulations imposed by institutions. A positive (negative) ϕ corresponds to a net positive (negative) influence on the agents' behaviour, making them more cooperative (defective). We use a small but positive external field, which represents a liberal society that is promoting cooperation on collective issues.

The agent's resistance to change is represented by w_i . Experimentation showed that our results do not qualitatively depend on w_i 's distribution across agents, so we assume them to be the same for all agents. The interaction strength between agents i and j is w_{ij} , which represents a combination of psycho-social factors such as relatedness, charisma, and cultural differences. We take this parameter to be normally distributed, since it reflects a combination of a large number of random influences. To accommodate the many highly variable instantaneous influences on social interactions, such as mood, location and medium of communication, we screen the interactions with a uniformly distributed stochastic noise term, $-r \leq \xi \leq r$, where r is the width of the distribution.

The factor, $|b_i - b_j|$ in Eq. (4) is introduced to ensure that the final opinion of an agent after an interaction lies inside or at least near the range of opinions of the interacting agents. This reflects the fact that agents can only change to other opinions that are actually presented to them. Consequently, meaningful changes in an agent's opinion primarily occur in exchanges with agents not already sharing their opinion. Conversely, any interaction between two already agreeing agents will only work preserve the agent's present opinions.

Since we are concerned with building collective action from a state of inaction, our simulations begin with network states in which most agents defect. Except where specified otherwise, they begin from an uncorrelated Gaussian distribution of agent behaviours with mean -0.25 and standard deviation $= 0.15$.

We use the Holme–Kim algorithm²⁹ to construct our clustered scale-free networks. We also analyse simple grid networks for pedagogical purposes. Unless otherwise stated, all of our networks consist of $33^2 = 1089$ nodes, with average degree $\langle k \rangle = 16$. The average degree is obtained as a combination of the degree of the growth nodes added by the Holme–Kim algorithm ($k_0 = 8$), as well as the accompanying clustering parameter ($k_t = 3.5$). We further use external field strength $\phi = 0.05$, agent self weight $w_i = 0.6$ and an inter-agent weight normal distribution with mean 0.5 and standard deviation 0.15 . Furthermore, the uniform interaction noise term ξ is bounded by $r = 0.1$. Under these parameters, which we call standard, the network always converges to a fully cooperative state, allowing us to focus our investigation on the dynamics of this process. All of the above parameters are summarized in Table 1.

The stubbornness and friendship parameter values were loosely chosen in the mid range because with these parameters the model behaves in a way that is qualitatively similar to what is observed in the real world: people can influence each other but are generally reluctant to change their opinions, and this varies widely due to many parameters³⁰. We propose that refining this aspect of the model is an interesting avenue for future work, based on data from ongoing empirical work on opinion change.

We study the collective dynamics across multiple simultaneously evolving networks. By obtaining data about the average cooperativity and standard deviation of cooperativity of agents in the networks, we can learn about what factors impact the network evolution and how. Throughout this paper, whenever we refer to the *standard deviation*, we are referring to the average of standard deviations across the individual networks. The average state contains quantitative information about the evolution of the system, such as in which direction it is evolving, and at which rate. Combining this with the standard deviation, and prior knowledge about the network topology we can draw qualitative conclusions about how the network is evolving, and what mechanisms drive this evolution.

All simulations were run using Python 3.7.2 with the Networkx 2.3 (for network support) and Community 1.0.0b1 (for the Louvain algorithm³¹) packages. Native parallelization routines were provided by the multiprocessing package.

Wherever we refer to the convergence rate in quantitative terms, we use the derivative at average cooperativity zero as an indication of the this rate. The derivative itself was found by linear regression around the data point closest to the root of the curve, we used a 10 point regression to safely eliminate numerical instabilities.

For a more thorough explanation of the model parameter values, as well as a the model's dependence on these, we refer to the Supplementary Material.

Received: 10 December 2020; Accepted: 17 March 2021

Published online: 28 April 2021

References

1. Ringsmuth, A. K., Lade, S. J. & Schlüter, M. Cross-scale cooperation enables sustainable use of a common-pool resource. *Proc. R. Soc. B* **286**, 20191943 (2019).
2. Min, Y., Du, Y. & Jin, C. The effect of link rewiring on a coevolutionary common pool resource game. *Physica A* **512**, 935–944 (2018).
3. Chen, X. & Szolnoki, A. Sanctioning and inspection for governing the commons in a feedback-evolving game. *PLoS Comput. Biol.* **14**(7), e1006347 (2018).
4. Szolnoki, A. & Chen, X. Environmental feedback drives cooperation in spatial social dilemmas. *Europhys. Lett.* **120**(5), 58001 (2017).
5. Dobay, A., Bagheri, H. C., Messina, A., Kümmerli, R. & Rankin, D. J. Interaction effects of cell diffusion, cell density and public goods properties on the evolution of cooperation in digital microbes. *J. Evol. Biol.* **27**, 1869–1877 (2014).
6. Tavoni, A., Schlüter, M. & Levin, S. The survival of the conformist Social pressure and renewable resource management. *J. Theor. Biol.* **299**, 152–161 (2012).
7. Scheffer, M. *et al.* Anticipating critical transitions. *Science* **338**, 344–348 (2012).
8. Buck, S. *The Global Commons: An Introduction* (Island Press, 2012).
9. Ostrom, E. A multi-scale approach to coping with climate change and other collective action problems. *Solut. J.* **1**(2), 27–36 (2010).
10. Heckathorn, D. The dynamics and dilemmas of collective action. *Am. Sociol. Rev.* **61**(2), 250–277 (1996).
11. Perc, M. *et al.* Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51 (2017).
12. Lorenz, J. Continuous opinion dynamics under bounded confidence: A survey. *Int. J. Mod. Phys. C* **18**, 18191838 (2007).
13. Deffuant, G., Neau, D., Amblard, F. & Weisbuch, G. Mixing beliefs among interacting agents. *Adv. Complex Syst.* **3**, 87–98 (2000).
14. Hegselmann, R. *et al.* Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artif. Soc. Soc. Simulat.* **5**(3), 1 (2002).
15. A. Das, S. Gollapudi, K. Munagala. 'Modeling opinion dynamics in social networks'. In *Proc. 7th ACM International Conference on Web Search and Data Mining* 403–412 (2014).
16. Ghaderi, J. & Srikant, R. Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica* **50**(12), 3209–3215 (2014).
17. Banisch, S. & Araújo, T. On the empirical relevance of the transient in opinion models. *Phys. Lett. A* **374**(31–32), 3197–3200 (2010).
18. De, A., Bhattacharya, S., Bhattacharya, P., Ganguly, N. & Chakrabarti, S. Learning linear influence models in social networks from transient opinion dynamics. *ACM Trans. Web* **13**(3), 1–33 (2019).
19. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**(2), 591 (2009).
20. Milgram, S. The small world problem. *Psychol. Today* **2**(1), 60–67 (1967).
21. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**(6684), 440 (1998).
22. Ugander, J., Karrer, B., Backstrom, L. & Marlow, C. The anatomy of the facebook social graph. Preprint at <https://arxiv.org/abs/1111.4503> (2011).
23. Karinthy, F. 'Chain-links'. In *Everything is Different* (1929).
24. Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**(5439), 509–512 (1999).
25. Fotouhi, B. N., Momeni, B., Allen, B. & Nowak, M. A. Conjoining uncooperative societies facilitates evolution of cooperation. *Nat. Hum. Behav.* **2**(7), 492–499 (2018).
26. Pham, T. M., Kondor, I., Hanel, R. & Thurner, S. The effect of social balance on social fragmentation. *J. R. Soc. Interface* **17**, 20200752 (2020).
27. Quattrociocchi, W., Caldarelli, G. & Scala, A. Opinion dynamics on interacting networks: Media competition and social influence. *Sci. Rep.* **4**, 4938 (2014).
28. Li, T. & Zhu, H. Effect of the media on the opinion dynamics in online social networks. *Physica A* **551**, 124117 (2020).
29. Holme, P. & Kim, B. J. Growing scale-free networks with tunable clustering. *Phys. Rev. E* **65**(2), 026107 (2002).
30. Mallinson, D. J. & Hatemi, P. K. The effects of information and social conformity on opinion change. *PLoS ONE* **13**(5), e0196600 (2018).
31. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**(10), 10008 (2008).

Acknowledgements

This work was supported by Swedish Research Council (Vetenskapsrådet) Grant 2015-0496, European Research Council (ERC) Grant 682472-MUSES, Austrian Science Fund (FWF) Grant P29252 and Austrian Research Agency (FFG) Grant 882184. We acknowledge Signe Savén for ideas that inspired this research, and Tuan Pham for enlightening discussions. We also extend our gratitude to Effective Altruism for promoting collective action on important issues.

Author contributions

D.A., A.S.d.W. and A.K.R. conceived the study. All authors formulated the model. S.B. wrote the simulation code base and produced preliminary results. D.A. further developed the code and produced the results. A.S.d.W. and A.K.R. advised on interpreting the results. D.A., A.S.d.W. and A.K.R. wrote the paper.

Funding

Open access funding provided by Stockholm University.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-87109-x>.

Correspondence and requests for materials should be addressed to A.S.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021