Social norms and groups structure safe operating spaces in renewable resource use in a social–ecological multi-layer network model

Max Bechthold[1,2,7], **Wolfram Barfuss**[3,1], **André Butz**[4], **Jannes Breier**[1,7], **Sara M. Constantino**[5,10], **Jobst Heitzig**[6], **Luana Schwarz**[1,7,9], **Sanam N. Vardag**[4], **and Jonathan F. Donges**[1,8,7]

[1]Earth System Analysis & Earth Resilience Science Unit, Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Telegrafenberg A31, 14473 Potsdam, Germany
[2]Institute of Physics and Astronomy, Potsdam University, Potsdam, Germany
[3]Center for Development Research (ZEF), University of Bonn, Genscherallee 3, 53113 Bonn, Germany
[4]Institute of Environmental Physics, University of Heidelberg, Im Neuenheimer Feld 229, 69120 Heidelberg, Germany
[5]School of Public Policy and Urban Affairs, Northeastern University Boston, MA 02115, USA
[6]Complexity Science Department, Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Telegrafenberg A31, 14473 Potsdam, Germany
[7]Department Integrative Earth System Science, Max Planck Institute of Geoanthropology, Kahlaische Strasse 10, 07745 Jena, Germany
[8]Stockholm Resilience Centre, Stockholm University, Frescativägen 8, 106 91 Stockholm, Sweden
[9]Institute of Environmental Science and Geography, Potsdam University, Potsdam, Germany
[10]Department of Psychology, Northeastern University Boston, MA 02115, USA TSI

Correspondence: Max Bechthold (maxbecht@pik-potsdam.de)

Abstract. Social norms are a key socio-cultural driver of human behaviour and have been identified as a central process in potential social tipping dynamics. They play a central role in governance and thus represent a possible intervention point for collective action problems in the Anthropocene, such as natural resource management. A detailed modelling framework for social norm change is needed to capture the dynamics of human societies and their feedback interactions with the natural environment. To date, resource use models often incorporate social norms in an oversimplified manner, as a robust and detailed coupled social–ecological model, scaling from the local to the global world–Earth scale, is lacking. Here we present a multi-level network framework with a complex contagion process for modelling the dynamics of descriptive and injunctive social norms. The framework is complemented by social groups and their attitudes, which can significantly influence the adoption of social norms. We integrate the modelling concept of norms together with an additional individual learning component into a model of coupled social–ecological dynamics with a closed feedback loop, implemented in the copan:CORE framework for world–Earth modelling. We find that norms generally bifurcate the behaviour space into two extreme states: one sustainable and one unsustainable. Reaching a sustainable (i.e. safe) state becomes more likely with low thresholds of conforming to sustainable norms, as well as lower consideration rates of own resource harvesting success. Modelling both descriptive and injunctive norms independently and dynamically introduces additional intermediate states, e.g. when there are countervailing norms. The shape of the bifurcation depends on the number of groups and members and thus on the social network topology. Where groups are very inert in changing their attitudes and thus consistently convey the same norm, multiple stable basins for sustainability levels are found. Groups influence the dynamics by facilitating or inhibiting the contagion of sustainable behaviour by communicating their norms. The success of a generic social norm intervention

is also found to be highly dependent on the group topology. Our findings suggest that explicitly modelling social norm processes together with social groups enriches the dynamics of social–ecological models and determines safe operating spaces. Consequently, both should be taken into account when representing human behaviour in coupled world–Earth models.

1   Introduction

The Anthropocene comes with accelerating entanglement, feedback interactions, and coevolution between the Earth system and the economic and socio-cultural Anthroposphere (Schellnhuber, 1999; Crutzen, 2002; Steffen et al., 2015; Lenton and Latour, 2018). Current assessment models of global change include only a limited number of these feedback interactions, hampering solution-oriented research (Verburg et al., 2015).

To capture the global coevolution of the entangled social "world" and the bio-geo-physical Earth system (Donges et al., 2020), it is necessary to close the loop and integrate dynamics of human systems and Earth system models (Donges et al., 2017a, b; Beckage et al., 2018; Calvin and Bond-Lamberty, 2018; Beckage et al., 2020; Steffen et al., 2020). *World–Earth modelling* (WEM) is the attempt to explicitly account for human–Earth system interactions (Nitzbon et al., 2017; Donges et al., 2020, 2021; Anderies et al., 2023).

1.1   Social norms & social groups

Social norms are a main socio-cultural factor (Opp, 2001; Legros and Cislaghi, 2019) of which WEMs lack a detailed representation (Donges et al., 2020). A major augmentation of (world–)Earth system assessments that researchers propose is therefore the inclusion of social norms and their dynamic coevolution with changes in the ecological environment (Kinzig et al., 2013; Donges et al., 2017b; Bury et al., 2019).

There is an extensive literature on social norms in many disciplines, such as sociology, social psychology, philosophy, anthropology, law, and economics (Young, 2015; Legros and Cislaghi, 2019; Gelfand et al., 2024). Especially the emergence of norms (e.g. Axelrod, 1986; Opp, 2001; Voss, 2005), the evolution of norms (e.g. Ehrlich and Levin, 2005; Young, 2015), the diffusion of norms (e.g. Epstein, 2001; Centola, 2015), and their influence on individual behaviours and cooperation (e.g. Ostrom, 2000) have been a focus of investigation. In the context of agent-based social simulation, the modelling of social norms has a long history (Schelling, 1971, 1978) with models varying in complexity, depending on the social setting and its relevant processes, for example (Bianchi and Squazzoni, 2015). However, much about social norms is not yet fully understood, such as how they could be best used for addressing global challenges of the Anthropocene, for which they may offer solutions (Ehrlich and Levin, 2005; Nyborg et al., 2016; Andrighetto and Vriens, 2022; Gelfand et al., 2024).

Thus, modelling and investigating social norms and their feedback interactions in world–Earth systems as a tool to investigate such global challenges is of manifold interest. Four reasons can be identified in particular.

First, since social norms govern human behaviour (Cialdini et al., 1990) and interactions between individuals (Young, 2015; Bicchieri, 2017), they are part of many major conceptualisations of human behaviour (Constantino et al., 2021a), such as, for example, the Theory of Planned Behaviour (Ajzen, 1991; Beckage et al., 2018; Ceschi et al., 2021). Hence, their inclusion can be an important step towards better modelling of socio-cultural processes and human behaviour in holistic WEMs and possible gains in assessing highly entangled systems that come with it (Beckage et al., 2020).

Second, social norms are expected to play a critical role in collective action problems (Ostrom, 2000; Nyborg et al., 2016; Constantino et al., 2022), where joint action of individuals to reach a common and beneficial goal is hindered by conflicting individual incentives (Olson, 1971). Major examples of such problems are climate change (Nyborg et al., 2016; Bak-Coleman et al., 2021) and the management of common resources, such as water conservation (Janssen, 2017; Castilla-Rho et al., 2017).

Third, social norms are suggested to play a pivotal role in overarching politics and governance (Finnemore and Sikkink, 1998; Biermann et al., 2010) and, as such, in designing interventions and policies as a response to climate change (Nyborg et al., 2016; Falk et al., 2021; Constantino et al., 2022).

Fourth, norms may be an important social tipping element (Schellnhuber, 2009; Andreoni et al., 2021) for rapid decarbonisation and sustainability transformations (Nyborg et al., 2016; Otto et al., 2020; Winkelmann et al., 2022). The interplay between individual behaviours and norms on different timescales could be a key for explaining abrupt regime shifts that alter the socio-cultural landscape (Ehrlich and Levin, 2005).

Summarising, social norms may hold the key to avoid drastic adverse global impacts due to climate change (Ehrlich and Levin, 2005), but can also drive collapse via unsustainable status-quo norms (Constantino et al., 2022).

We additionally include social groups for the socio-cultural modelling, which is motivated by the strong connection of groups with social norms (Tomasello, 2019). So-

cial groups commonly include norms that influence interactions and actions of their members (Homans, 2017; Sherif and Sherif, 1965; Forsyth, 2018; APA, 2023). Norms can differ between groups, where adherence to group norms can strengthen membership and group identity (Bernhard et al., 2006). The propensity to join groups is an important characteristic of human behaviour. Processes that unfold within and through these groups indelibly influence group members and society (Forsyth, 2018). Societal transformation, such as a shift to a net-zero emissions economy, might threaten certain group interests or identities (Constantino et al., 2022). Thus, social groups are an important extension for WEMs (Donges et al., 2020).

1.2 Coupled natural resource management

In this study, we focus on natural resource management. Natural resource management provides examples of systems with high entanglement between the social and the ecological compartment (Schlüter et al., 2012) that are intended to be analysed with WEMs (Donges et al., 2020). Resource extraction is intensifying CO_2 emissions (Liu et al., 2022) and is posing a major threat to planetary boundaries (Rockström et al., 2009), for example in the form of the deforestation of the Amazon rainforest (Malhi et al., 2008). Models with coupled dynamics between social and ecological processes can contribute to a better understanding of the sustainable management and adaptation to global change of such natural resource management systems (Schlüter et al., 2012). In particular, coupling models of social norms with biophysical models can be crucial in evaluating pathways for sustainability in natural resource use models (Janssen, 2017; Castilla-Rho et al., 2017). At the same time, social norms also play a very important role in empirically studied, real-life resource use situations (Ostrom, 2000), for example mismanagement of fishing regulations in Norway (Maurstad, 2000). Thus, we here develop a dynamical social norm framework in an agent-based model and employ it in the setting of a coupled model of resource use to comparatively showcase the modelling opportunities and its dynamics. This implementation can be used to extend world–Earth models with a detailed normative dimension.

Human–environment interactions and the individual resource stocks are modelled following Wiedermann et al. (2015); i.e. agents harvest a logistic resource with a binary effort level, leading the resource stock to either collapse or reach a sustainable level. The resource stock is thus used as a measure of sustainability and only a sustainable state considered a "safe operating space" (Heitzig et al., 2016).

Here, human–environment interactions are conceptually assumed to happen in a private good or private-pool setting; that is, each agent has its individual resource stock only accessible by that agent (Barfuss et al., 2017). This can be thought to represent a setting in which individual foresters manage one piece of forest. Thus, information on the harvesting behaviour of other agents is purely retrieved via social interaction. While this modelling decision has the caveat of not allowing one to include common mechanics of cooperation in common-pool resource models (Tavoni et al., 2012; Tu et al., 2024), it puts the focus of the model specifically on processes in the socio-cultural domain, highlighting group- and norm-influenced spreading of behaviour, i.e. the social norm framework. It also keeps results comparable to former work with the same setting (Wiedermann et al., 2015; Barfuss et al., 2017; Geier et al., 2019). The exploration of a common-pool setting is left for future work.

1.3 Social norm modelling framework

Many models that couple social dynamics and resource dynamics include social norms as a uniform pressure on individuals' opinion formation (Barlow et al., 2014; Ali et al., 2015; Bauch et al., 2016; Sigdel et al., 2017; Thampi et al., 2018; Bury et al., 2019), as a parameter of the Theory of Planned Behaviour (Beckage et al., 2018), or as an optimal strategy to adhere to in an evolutionary game theory perspective (Tavoni et al., 2012; Lade et al., 2013; Schlüter et al., 2016; Tilman et al., 2016; Farahbakhsh et al., 2021). It has been found that the norm dynamics can be essential to the overall outcome of the resource models (Farahbakhsh et al., 2022), for example by enforcing a single strategy supporting one of two equilibria with either sustainable or disastrous outcomes for the resource stock (Satake et al., 2007; Lade et al., 2013; Bauch et al., 2016; Sigdel et al., 2017).

These approaches are not able to capture the full spectrum and dynamics of social norms in coupled systems as in reality, populations usually do not converge uniformly to one strategy (Bauch et al., 2016) and norms are typically said to have multiple equilibria, e.g. supporting multiple strategies (Young, 2015). It has been proposed that the impact of norms on human behaviour can only be usefully understood when making the distinction between descriptive and injunctive norms (Cialdini et al., 1990). Their understanding is crucial when designing interventions (Constantino et al., 2022) so as to prevent possible "boomerang effects" (Schultz et al., 2007). Still, there is either commonly only one of both types represented or they are represented as one overarching norm, lacking this distinction (Farahbakhsh et al., 2022).

For the social norm modelling framework, we therefore use a definition for social norms following Bicchieri (2017), which crucially includes this distinction between *descriptive norms* (a) and *injunctive norms* (b): a social norm is a rule of behaviour such that individuals prefer to conform to it on the condition that they believe that (a) most people in their reference network conform to it [...] and [they believe that] (b) [...] most people in their reference network believe they ought to conform to it [...]. The reference network denotes all individuals that are relevant for the decision to conform to a behaviour, i.e. adopt it due to normative pressure. This reference network can relate to relatives, friends, neighbours, or

unknown people in public situations, from single individuals to groups.

There have been few attempts (Suzuki and Iwasa, 2009; Nøstbakken, 2013; Lin et al., 2022) to represent both descriptive and injunctive norms as distinct processes with endogenously arising dynamics, i.e. arising from within the model processes, as compared to exogenous, prescribed parametric pressures. This holds in particular for injunctive norms.

In this study, we depart from here and intend to model social norms including both descriptive and injunctive norms, as two dynamic and endogenously arising processes, in contrast to former studies. These processes govern the complex contagion of behaviour; i.e. a node needs to be influenced by multiple neighbours to adopt, which is a robust assumption for social norms (Centola and Macy, 2007; Guilbeault et al., 2018). This implementation includes a threshold approach (Granovetter, 1978); i.e. a norm needs to surpass a certain majority to be considered by agents (Centola et al., 2005; Müller-Hansen et al., 2017). With our approach we try to strike the right balance between a novel representation of endogenous norms within groups and model tractability. For this we explicitly leave some features that are commonly modelled in agent-based models of social norms in resource applications, such as sanctioning (Tavoni et al., 2012; Nhim et al., 2019), to future work.

Parametric norm representations also do not relate to the pattern of social ties, i.e. network topology (Centola and Macy, 2007; Centola, 2010), or groups in particular (Centola, 2015), which can both crucially influence the spreading of social norms. Since social groups and social norms are strongly tied structures (Tomasello, 2019), social groups play an important role in the model. They do so by mediating the injunctive norm, a process which is laid out in detail in the following Sect. 2.3. The inclusion of social groups is achieved by expanding a "lower" social network layer of agents that represent informally connected individuals, with an "upper" layer of agents that represent formal social groups in a multi-layer network (Boccaletti et al., 2014), where membership is indicated by links between the network layers. This constitutes a basic social structure (Centola and Macy, 2007; Davis et al., 2015; Guilbeault et al., 2018; Newman, 2018), including variable topologies (i.e. number of groups, or shape and size of member bases of a group). By adding groups into the agent-based model, we aim to have norms reach multiple behavioural equilibria, extending the model dynamics beyond a single strategy outcome. We also intend to analyse how different group structures (number of groups etc.) can influence the success of a stylised social norm intervention, where groups are said to be pivotal in leading change (Constantino et al., 2022).

1.4 The Nexploit model

By integrating the social norm and groups modelling framework with the resource use application, we obtain a coupled model called *Nexploit* model (following the *Exploit* model research line).

To improve integrability of the norm framework and the coupled model into advanced WEMs, both are implemented in *copan:CORE*. The open world–Earth modelling framework copan:CORE enables flexible modelling with standard components for developing, composing, and analysing WEMs across the spectrum from stylised and aggregated to complex and spatially and socially highly-resolved model variants. Based on elementary entity types (grid cells, individuals, social systems), copan:CORE models can contain processes and feedbacks between them in different conceptual taxa: biophysical (e.g. resource growth), socio-metabolic (e.g. resource extraction), and socio-cultural (e.g. social learning) (Donges et al., 2021). The modular approach supports systematically comparing, testing, reusing, and exchanging components and their theoretical assumptions (Donges et al., 2020). In Fig. 1, we show an overview of the concepts, entity types, and process types from copan:CORE that are used in this model and their classification within the taxonomy. For a detailed description of copan:CORE, see Donges et al. (2020). With this work, we add the new entity type "Group" (Bechthold et al., 2024) to the framework. We also provide the Nexploit model and our social norm framework as a ready-to-use component for copan:CORE.

The model is constructed by three interacting components, each of them classified according to the taxonomy for structuring models for world–Earth system analysis (Donges et al., 2021) and each of them carrying one main conceptual process. The first component models a stylised ecosphere, with the growth of a renewable resource in a resource-limited environment, that belongs to the biophysical taxon (ENV). The second component models stylised human–environment interactions, i.e. harvesting of the corresponding resource, reflected in the socio-metabolic taxon (MET). The third component models a stylised anthroposphere, with social norms and individual learning determining harvesting behaviours, that belongs to the socio-cultural taxon (CUL).

These processes lead up to different process interactions and a closed interaction loop, which are shown in Fig. 2, a conceptual representation of the model.

The model is designed with the intention to explore the coevolution and entanglement of a stylised socio-cultural sphere and a stylised ecosphere, mediated by stylised metabolic processes. The nearest goal of the model is to determine under which constellations of social norms and harvest effort the collection of resources collapses from overexploitation and to explore whether generalisable conclusions can be drawn from this. In particular, we seek to analyse the novel social norm framework, including social groups, and its influence on the system dynamics. Due to the complexity of human behaviour and social dynamics (Carpenter et al., 2019), the task is not to predict future pathways, but to eval-
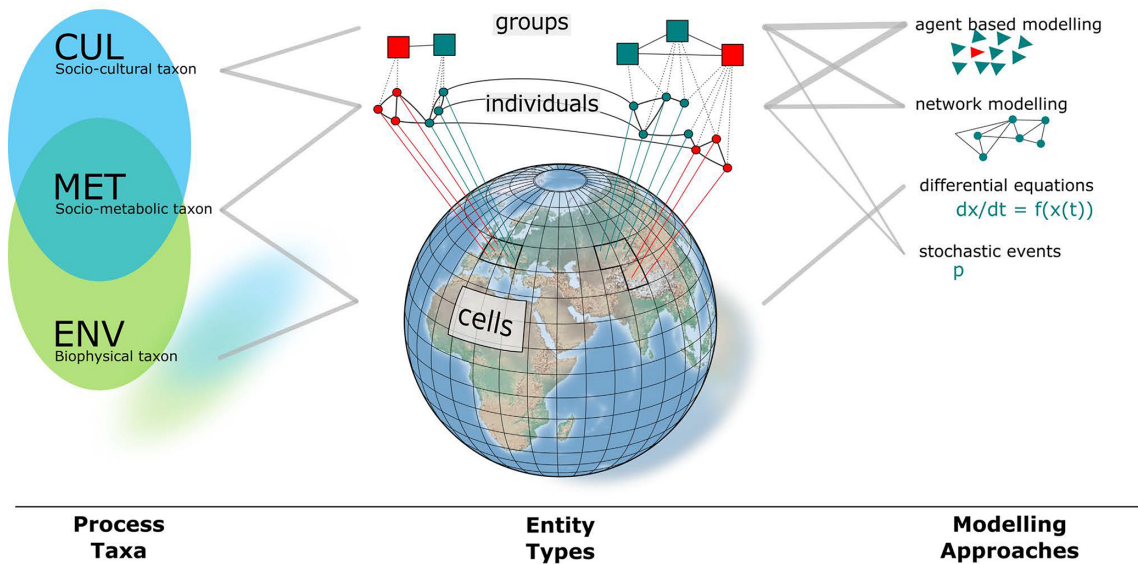
Figure 1. Overview of the copan:CORE open world–Earth modelling framework, referring to the parts relevant to the present work. (middle) The entities in copan:CORE models are formed via entity types (e.g. cells, individuals, groups). The novel group CEI entity is indicated by blue spheres. All processes belong to a process taxon (left); i.e. resource growth belongs to the biophysical taxon. Processes are distinguished by their formal process types (right). It is possible to freely combine entity types, process types, and process taxa (grey lines), where thick grey lines illustrate combinations used in this model. Adapted and modified from Donges et al. (2020).

Figure 2. Classification of the relevant Nexploit model processes, ordered by owning process taxon (row) and affected process taxon (column). Arrows indicate how processes feed into other processes, forming a loop between biophysical, socio-metabolic and socio-cultural dynamics. ENV → ENV: the generation or growth of the resource is a purely environmental process, but influences the possible harvesting yield. MET → CUL: how successful an individual harvests generates utility, which influences the social norms and individual learning process. CUL → CUL: on a purely socio-cultural level, social norms and individual learning take place. CUL → MET: depending on social processes, an individual selects higher or lower efforts, influencing the harvest. MET → ENV: depending on the harvesting behaviour of the individual, the resource generation is more or less restricted by extraction.

uate possible interventions to reach desirable states of the "whole Earth system" in the long run.

The coupled model has three, to our knowledge unique, key aspects that are seminal for future world–Earth modelling: (a) the explicit inclusion of endogenous dynamic descriptive and injunctive norms in a coupled model of resource use, through (b) the inclusion of social groups in (c) a multilayer network. This integration enables us to investigate the influence of social norms and social norm interventions in an exemplary collective action problem in more detail and taking into account the distinction between descriptive and injunctive norms, social groups, or the network topology.

With the model, we find that two distinct norm processes strongly influence the resource dynamics and structure the state space of sustainable and unsustainable model outcomes. Additionally to a possible sharp bifurcation of the system into two extreme states, one sustainable and one unsustainable, it introduces equilibria of intermediate sustainability, e.g. where one norm process is strong and the other is not. The presence of groups that send the same normative message over very long timescales, i.e. being sustainable, also allows the model to reach intermediate levels of sustainability. The parameters that control these outcomes are the thresholds for norm uptake from the social norm framework as

well as the parameters that determine the group topology, i.e. number of groups and average group size. These parameters also determine the effectiveness of "social-norm interventions" in the model.

2   Model description

In the Nexploit model, private-pool resources with logistic growth dynamics are harvested either sustainably or unsustainably by agents, following the resource dynamics of a well-investigated, social–ecological network model, the Exploit model and its extensions (Wiedermann et al., 2015; Barfuss et al., 2017; Geier et al., 2019). The social norm framework then, through social norms and groups, influences the choice of sustainable or unsustainable harvesting effort, where an unsustainable effort leads to the collapse of the private resource. Social interactions thus govern the state of the environment (the resource availability). The harvesting then closes the loop: the yield, which depends on the state of the environment, drives an individual learning process. This is introduced additionally to the social norm dynamics, which can be broadly understood as a social learning process instead (Bandura, 1977; Reed et al., 2010). A higher yield reduces the probability of switching behaviour, similar to a win–stay, lose–shift strategy (Nowak and Sigmund, 1993), which could be interpreted as a change-inhibiting process of social inertia (Brulle and Norgaard, 2019) as well. Therefore, the state of the environment also feeds back into the social dynamics and vice versa.

Individual learning can be seen as a dynamic account of classic utility-maximising behaviour (Barfuss, 2022), in which an agent favours the option with the highest utility according to its preferences. This behaviour, however, is prone to collective action problems (Ostrom, 2000; Barfuss et al., 2020), in which social norms play a crucial role (Nyborg et al., 2016; Janssen, 2017; Castilla-Rho et al., 2017; Constantino et al., 2022).

In Fig. 3, an overview of the multi-layer network structure, all processes, and their classification into the taxonomy scheme in a visual representation can be seen.

From the entity types that are provided by copan:CORE, the "Individual" entity type is used to represent the agents. Each agent a_i is assigned to one "Cell" entity type c_i and one only. Therefore, the same index i always denotes a cell and an agent that belong together. Then, the number of cells N_c is equal to the number of agents N_a and $N_c = N_a = N$. The novel "Group" entity type, is used to represent groups g_k of agents.

2.1   Biophysical component: resource growth

The logistic growth model, which models growth in a resource-limited environment, has become a standard for stylised resource growth representations in social–ecological models (Farahbakhsh et al., 2022). For the case of a renewable resource that is harvested, the open-access fishery model is a well-known example (Perman, 2003). Building upon this, a simplistic model for the resource dynamics in N local stocks s_i (without harvesting yet) can be deduced (Wiedermann et al., 2015):

$$\frac{\mathrm{d}s_i(t)}{\mathrm{d}t} = \lambda_i s_i(t)\left(1 - \frac{s_i(t)}{K_i}\right), \tag{1}$$

with $\lambda_i > 0$ the growth rates and $K_i > 0$ the bounding maximum capacities for $i = 1, \ldots, N$.

Here λ_i and K_i are assumed to be the same for all cells, i.e. there is homogeneous resource distribution, hence dropping the index and obtaining $\lambda_i = \lambda$ and $K_i = K$, for all $i = 1, \ldots, N$. Then, λ and K are set to unity, $\lambda = 1$, $K = 1$, to obtain a dimensionless measure of time and the stocks, with $0 \leq s_i \leq 1$.

2.2   Social-metabolic component: harvesting

The social-metabolic component represents the harvesting of the local resource by the individual agents through a harvest term

$$h_i(t) = s_i(t)E_i(t), \tag{2}$$

where E_i denotes the effort level of harvest of an agent. This harvest term is then subtracted from the resource stock of the cells, leading to (Wiedermann et al., 2015; Barfuss et al., 2017; Geier et al., 2019):

$$\frac{\mathrm{d}s_i(t)}{\mathrm{d}t} = s_i(t)(1 - s_i(t)) - h_i(t). \tag{3}$$

The harvest term will also play a role in the social dynamics and builds the bridge between the biophysical and the sociocultural component. Equation (3) fully describes the applied resource and harvesting dynamics of the system. From now on, the explicit time dependency of the stocks and efforts will be omitted.

The effort level E_i of an agent a_i is encoded by its behaviour B_i, which can be sustainable, $B_i = 1$, or unsustainable $B_i = 0$ in a binary way.

$$E_i = \begin{cases} \frac{1}{2} & \text{if } B_i = 1, \\ \frac{3}{2} & \text{if } B_i = 0, \end{cases} \tag{4}$$

is chosen for the low and high effort levels, such that the sustainable behaviour coincides with the maximum sustainable yield (Wiedermann et al., 2015; Barfuss et al., 2017) and such that $h_{\min} = 0$ and $h_{\max} = 1.5$. This results in a stable fixed point of Eq. (3) at $s_{0,-} = 0$, for a high effort level. A high effort level leads to a state of complete resource depletion, elucidating why this effort level (or behaviour) is considered unsustainable. For a low effort level (or sustainable behaviour), a stable fixed point with $s_{0,+} = 0.5$ is found, which is considered sustainable.

It follows that the behaviour B_i and the resulting effort level E_i are the model variables that determine the state of a system to be sustainable or unsustainable.
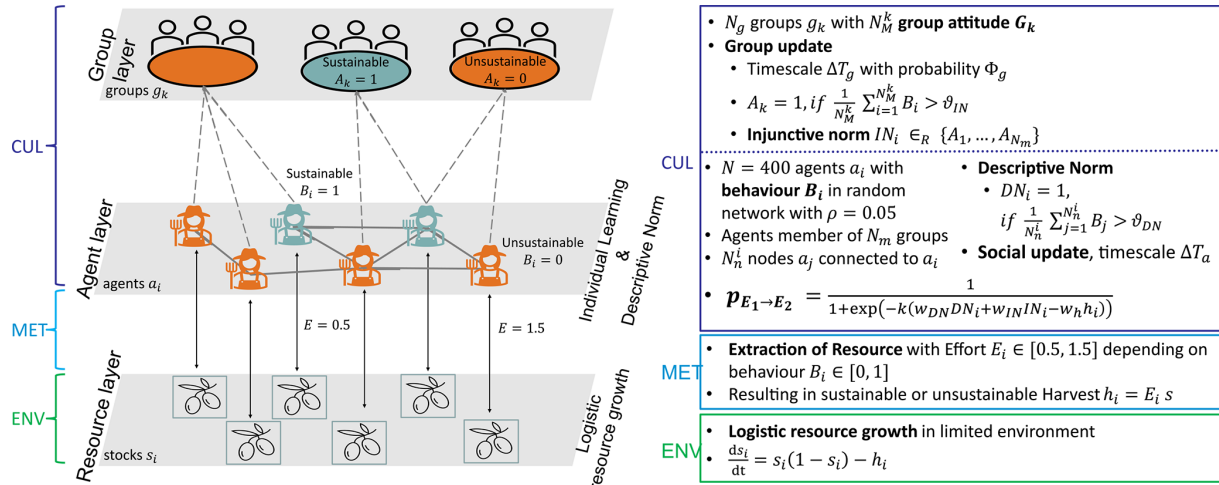
Figure 3. Overview of the Nexploit model, its feedback interactions, and the classification of its subcomponents into the copan:CORE framework. The resource growth and extraction dynamics (ENV and MET) follow well-established models. The socio-cultural dynamics (CUL) represent the social learning of either sustainable or unsustainable strategies under the influence of social norms, i.e. descriptive and injunctive norms. The descriptive norm is modelled on the agent layer, where agents "see" what the majority of neighbouring nodes do, while the injunctive norm is represented through the influence of groups.

2.3 Socio-cultural component: social norm framework

The socio-cultural component contains the social norm framework, where agents are subject to social learning (Bandura, 1977; Reed et al., 2010) of a binary behaviour (sustainable or unsustainable) in a multi-layer network (Boccaletti et al., 2014) that represents a basic social structure (Centola and Macy, 2007; Davis et al., 2015; Guilbeault et al., 2018; Newman, 2018), including groups on a second network layer. The normative behavioural spreading is modelled as a complex contagion (Granovetter, 1978; Centola and Macy, 2007) process, a solid assumption for social norms (Guilbeault et al., 2018), where the behaviour of majorities influences the behaviour of individuals (Müller-Hansen et al., 2017).

We consider an individual learning component as an additional baseline factor to the social learning through norms. Here, agents are less likely to switch their harvesting effort when they are currently successful, in line with the famous game-theoretic strategy Pavlov, also known as win–stay, lose–shift (Nowak and Sigmund, 1993). This component connects the state of the resource (environmental taxon) to the agents' decision-making (cultural taxon) (Fig. 2).

The normative social learning and the individual learning are then integrated into a decision-making function, where on an average timescale ΔT_a, agents consider updating their effort with some probability p, depending on how they are influenced by social norms, groups, and harvest.

With this approach we include a dynamic, endogenous account for the social influence of descriptive and injunctive norms through different groups and analyse its influence on a private-pool resource use, together with a rather evolutionary game-theoretic representation of individual learning. It

should be noted that our model aims to improve social norm representation in coupled resource-use models in this regard, not to give an exhaustive representation of all possible social norm effects. We exclude some elsewhere commonly included mechanisms of social norms and commons dilemma frameworks for the sake of simplicity of the agents and a lean and tractable model. Such mechanisms include, for example, cooperation (Tavoni et al., 2012; Tu et al., 2024), norm conformation (McDonald and Crandall, 2015), norm internalisation (Elsenbroich and Gilbert, 2015), norm sanctioning (Opp, 2001; Tavoni et al., 2012; Nhim et al., 2019), and further group effects (Spears, 2021). We will discuss the lack of such mechanisms in Sect. 4. Analysing the general dynamics of the novel model in this work will facilitate the inclusion of these mechanisms for a more integrated representation of social norms in future work.

2.3.1 Agent layer

The first layer, the agent layer $G_a = (\mathcal{V}_a, \mathcal{E}_a)$, represents a set $\mathcal{V}_a$ of $N_a = N_c = N$ agents that are connected through edges $\mathcal{E}_a$. These connections can be thought as connections in informal groups, such as being neighbours or having business relations, for example. Most importantly, an agent is able to "see" the behaviour of a connected node.

At the agent level, the network is modelled by the well-studied Erdős-Renyi random graph (Erdős and Rényi, 1960) with a link density $\rho = 0.05$. This network type is chosen according to earlier Exploit works for the network topology, so as to keep the coupled model comparable to these works (Wiedermann et al., 2015; Barfuss et al., 2017; Geier et al., 2019). Of course, this network type can be exchanged

for more detailed applications with a known social structure. Also, the number of agents (and cells) is set according to earlier studies, with $N_a = N_c = N = 400$. Agents are randomly initialised in a Bernoulli process by initially choosing one of the two behaviours with a probability of $p = 0.5$. This leads to a Binomial distribution of the behaviours in the beginning, with an expectation value for sustainable individuals of $\mu_{init} = N \cdot p = 200$ TS2.

2.3.2 Descriptive norm

The descriptive norm DN is thought to emerge from the direct interactions between agents, where the reference network of an agent a_i is thought to be all agents a_j it is connected to. An agent considers the descriptive norm DN_i that acts upon it to be either a sustainable (1) or an unsustainable (0) pattern of behaviour, depending on whether the mean of the behaviours B_j of its neighbours in the network N_n^i exceeds a certain threshold θ_{DN}: TS3

$$
DN_i \begin{cases} = 1 & \text{if } \frac{1}{N_n^i}\sum_j^{N_n^i} B_j > \theta_{DN}, \\ \in_R \{0,1\} & \text{if } \frac{1}{N_n^i}\sum_j^{N_n^i} B_j = \theta_{DN}, \\ \in_R \{0,1\} & \text{if } \frac{1}{N_n^i}\sum_j^{N_n^i} B_j = \theta_{DN}, \\ = 0 & \text{else,} \end{cases} \tag{5}
$$

where N_n^i is the number of neighbours, that is the number of agents that share an edge with agent a_i. A special case, where agents perform a random choice, denoted $\in_R$, when the mean behaviour equals the threshold is included to prevent network effects.

If we imagine the agents to be foresters, the descriptive norm forms through the interaction of one worker with, for example, neighbours whose harvesting behaviour could be directly observed on a day-to-day basis, for example due to adjacency or proximity of strips of forest.

2.3.3 Group layer

The second layer, the group layer $G_g = (\mathcal{V}_g, \mathcal{E}_g)$, represents a set $\mathcal{V}_g$ of N_g social groups g_k that can be connected through edges $\mathcal{E}_g$. Since there is no direct interaction between groups here, $\mathcal{E}_g = \{\}$.

We define a social group as consisting of two or more member individuals who understand themselves as part of the same social category (Turner, 2010; Geschke and Frindte, 2016). There is often a perception of unity inside the group and a perception of unity from outside the group (Forsyth, 2018), such that groups are perceived to be one actor. This motivates the fact that groups are being represented as independent agents in the model through a separate entity type in copan:CORE.

Group membership is represented by the interlayer connections $\mathcal{E}_{ag}$ between the aforementioned agent layer G_a and the group layer G_g. If an edge between an agent a_i and a group g_k exists, it represents the membership of the agent in the group. Through this membership, the behaviour of the agent influences the group attitude and vice versa. An agent has N_m^i memberships in different groups and a group has N_M^k members. In this model, all agents have the same number of memberships, therefore the index i is dropped and $N_m^i = N_m$ for readability.

The network is constructed by randomly connecting each agent to one group that it is not connected to yet, out of the pool of groups, N_m times. The parameter N_m therefore controls the shape of the random interlayer network $\mathcal{E}_{ag}$, with $N_m \in [0, N_g]$. For $N_m = 0$, agents are not connected to groups and hence not influenced by their dynamics. For $N_m = 1$, each agent is member of one group and for $N_m = N_g$, each agent is a member of all existing groups. Interactions in groups take place with relative permanence (Neidhardt, 2017), as people often remain members of groups for prolonged time periods. In the model, we consider stable groups as memberships do not change. An adaptive network, which could be used to model dynamic group memberships, was not included in this work, so as not to overshadow other model dynamics.

In reality, individuals can be members of different types of groups, informal and more formal ones (Tichy and Fombrun, 1979; Hunter, 2016), both of which can form reference networks for norms to their members. An informal group might be a group of colleagues, as are represented in the lower network layer as all the agents an agent is linked to.

The groups we model with the new entity type and in this second layer are meant to represent more structured formal groups, for example a more institutionalised workers union or association of foresters. Groups often have a shared group's viewpoint, strategy, stance or attitude (Forsyth, 2018), referred to as group attitude from here on. We represent this by the groups translating the behaviour of their members into an attitude or position towards said behaviour. Hence, each group has a group attitude $A_k \in [0, 1]$ regarding sustainable or unsustainable behaviour, where $A_k = 1$ denotes an approving attitude towards sustainable behaviour and $A_k = 0$ a non-approving one. In contrast to the descriptive norm, where an individual is thought to know the behaviour of all its reference network, in our group context, a member does not know the behaviours of all other members. It just knows the aggregate group's stance, which it receives as a message of what should be done from the group, without necessarily knowing what other members actually do. In real life this might correspond to a workers union or association that gives out recommendations on how to behave. Another example of such a process in a large scale formal group is the 2015 encyclical of pope Francis addressing climate change, resulting in varying levels of concern in U.S. Catholics (Li et al., 2016). Formal groups or institutions, that govern or educate a community can be very important for the diffusion of social norms (Constantino et al., 2021b).

Group attitudes have to be elaborated through member interactions (Forsyth, 2018) that often take place in regular intervals. This is mimicked by a regular event taking place in intervals of ΔT_g. With probability Φ_g, groups consider updating their group attitude. In this case, the mean of the behaviours B_i of member agents is calculated and if it surpasses a certain threshold θ_{IN}, the group attitude changes, similar to the way the descriptive norm is obtained:

$$A_k \begin{cases} = 1 & \text{if } \frac{1}{N_M^k}\sum_j^{N_M^k} B_j > \theta_{IN}, \\ \in_R \{0, 1\} & \text{if } \frac{1}{N_M^k}\sum_j^{N_M^k} B_j = \theta_{IN}, \\ = 0 & \text{else,} \end{cases} \quad (6)$$

where N_M^k is the number of members of group g_k. This group update is modelled through a step process of copan:CORE. Thus, the attitude is mainly determined by the threshold parameter θ_{IN}. The attitude in turn is used to determine the injunctive normative pressure on agents, such that θ_{IN} is said to represent the threshold for injunctive norm uptake. Similarly, the timescale ΔT_g is said to represent the characteristic timescale for the injunctive norm.

A special case can be constructed by not allowing groups to change their attitude. Groups then become a vessel of one attitude and therefore one norm that they keep sending at all times. In this case, they can be considered a "norm entity".

2.3.4 Injunctive norm

Since membership in a group is often connected to a degree of identification with the group and its predominant attitudes towards certain issues, the group attitude is thought to exert a normative expectation and pressure on its member individuals that we model as the injunctive norm IN.

In the model context, the design of the injunctive norm process can be illustrated going back to the example of the association of foresters: especially in larger formal groups that are work-related (e.g. a union of foresters), groups do not necessarily coincide with people someone directly works with, and thus the exact behaviour cannot be observed. Being a member of a group with a sustainable stance on resource use, a member might very well assume that it is expected from them to harvest sustainably, while they will not be able to actually observe whether all the other members also harvest sustainably at all times. Still, there will be pressure to conform due to the normative expectation that might be explicitly (e.g. in the statutes) or implicitly formulated by and within the group. In our case, groups might also represent non-work-related groups, such as friend circles or political parties, depending on the group size parameter.

Another example in which norms play a role is flight shame (Gössling et al., 2020). A member of a climate activist group might very well assume that it is expected from them not to fly often, while they will hardly be able to actually observe how often all the other members fly on a descrip-

tive level. The descriptive norm thus might be formulated by other peer groups, friends or family. Still, there will be pressure to conform to the injunctive norm of the activist group due to a strong sense of social identity and the normative expectation that might be explicitly or implicitly formulated by and within the group.

At each update an agent a_i considers the injunctive norm acting on it IN_i to be either a sustainable (1) or unsustainable (0) pattern of behaviour, by randomly choosing out of the set $S_{A_k}^i = \{A_1, \ldots, A_{N_m}\}$ of the group attitudes of its N_m group memberships:

$$IN_i \in_R S_{A_k}^i. \quad (7)$$

This random choice out of all group attitudes an agent is connected to, reproduces a situation where not all groups always have the same importance when decisions are taken. When the group attitudes are fixed, a group constantly sends the same injunctive norm. This mimics a fix injunctive norm or an injunctive norm entity in the socio-cultural domain.

2.3.5 Social update

The process through which an agent changes its effort is called social update. A social update is conducted as a random event, as pre-implemented in copan:CORE by the event process. Assuming the system to be at time t_0, the next updating interval τ is drawn from an exponential distribution:

$$p(\tau) = \frac{1}{\Delta T_a}\exp(-\tau/\Delta T_a), \quad (8)$$

a well-studied distribution for human interaction rates in queuing theory (Saaty, 1957), where ΔT_a gives the expected average time between two updates of agents. At the next updating time $t_1 = t_0 + \tau$, all agents are randomly ordered by an exponential distribution and then consider a social update with probability Φ_a. This updating probability controls the percentage of agents that enter the decision process for updating their state; i.e. $\Phi_a = 0.5$ implies that on average 50 % of agents enter the updating process. After a social update took place, a new updating interval is drawn and the next social update is conducted at $t_2 = t_1 + \tau$. The social update and the group update processes are independent of each other and their (average) waiting intervals, ΔT_a and ΔT_g, will be investigated thoroughly. Both are dimensionless measures here that have explanatory power in relative comparison only, although it is possible to implement data-informed timescales with copan:CORE.

When a social update is conducted for an agent, it will compare its current effort level E_j with the normatively encoded effort level of its relevant reference network. An agent will switch effort levels to E_k according to a sigmoid-shaped decision-making function (Traulsen et al., 2010), which in this case is modelled through a symmetric logistic function

(also called expit):

$$P\left(E_j \rightarrow E_k\right) = \frac{1}{1 + \exp(-kx)}. \tag{9}$$

This decision-making function is symmetric around 0. For $k = 0$, which controls the slope of the function, it approximates a flat line with $p = 0.5$ everywhere and for $k \rightarrow \infty$ it approximates a step function. For the special case of $k = 2$, the logistic function returns the same probability values as the tanh used in the original Exploit model, for same values of x. While x in the Exploit case is given as the difference in harvest between two agents, x is the sum of the two norms influencing an agent minus its harvesting success in this model:

$$x = w_{DN} \cdot DN' + w_{IN} \cdot IN' - w_h \cdot h', \tag{10}$$

where w_{DN}, w_{IN}, and w_h are the weights for the influence of each norm aspect (descriptive and injunctive) and the own harvesting success. The argument x is constrained to be $-1 \leq x \leq 1$, by adjusting the weights correspondingly. The weights can be further used to include effects where, for example, an injunctive norm is contrasted by a conflicting social norm, reducing its effect on behaviour (Smith et al., 2012; Staunton et al., 2014; Bonan et al., 2020b). Here, the weights will be assumed to be equal for all processes, so as to investigate the general dynamics of the framework for later use in copan:CORE. With relevant data, the weights could also be adjusted to according empirical studies.

To obtain the same probabilities to switch in both directions, the single terms of the sum are mapped accordingly. Hence, an unsustainable individual has the same probability to become sustainable if both norms are found to be sustainable, as a sustainable individual that is influenced by both unsustainable norms has, and so on. These mapped quantities are indicated by an apostrophe. If an agent switches its behaviour because of a social norm, it is said to conform to the norm.

The probabilistic nature of the decision-making function accounts for uncertainties or any (external) aspects of decision-making that the social norm framework and individual learning do not cover, such as involuntary choices, for example. It does not account for different reasons and ways that might lead agents to conform to a prevalent behaviour.

3 Results

For the analysis, each sub-process is assumed to have the same influence on decision-making; i.e. the weights in the function are set to $w_{IN} = 0.33$, $w_{DN} = 0.33$ and $w_h = 0.33$ such that, in sum, the focus lies on the social norm framework. This can be adjusted for studies with a different focus and a different context.

Since every node is described by one stock s_i and one behaviour B_i, the possibilities to visualise the dynamics of the system directly in phase space are limited. Instead, average

quantities are defined and investigated for their influence on the model, depending on the parameter choice. By keeping all parameters except for one or two fixed, their interplay can be analysed in parameter sweeps.

To reduce the parameter space that has to be investigated, some parameters are kept constant over the course of testing. They are chosen according to earlier studies with models from the Exploit family (Wiedermann et al., 2015; Barfuss et al., 2017; Geier et al., 2019): the number of agents $N = 400$, the two effort levels $E_s = 0.5$ and $E_u = 1.5$, and the average link density in the agent network $\rho = 0.05$. From further analysis, parameter choices $\Phi_a = 0.25$, $\Phi_g = 0.25$, and $k = 3$ are deduced, where such update probabilities Φ_a and Φ_g ensure smooth runs without jumps and $k = 3$ results in continuous switching probabilities between 5 % and 95 %, depending on the state of the norms and the current harvest of an agent (e.g. an agent that has $h = 0$ and is influenced by sustainable descriptive and injunctive norm aspects, has a probability to become sustainable of 95 %).

With the probabilistic decision-making function, the social component of the model includes a stochastic element. Therefore, an ensemble of runs is analysed for each parameter set and mean quantities are formed from the observables of interest. The number of simulation runs for each set is $n_r = 100$. From one (simulation) run, the mean over all N single stocks s_i and behaviours B_i is taken. The resulting, global stock fraction is then called S, where $S(t) = \frac{1}{N}\sum_{i=1}^{N} s_i(t)$. $S = 0$ would indicate that all resources have been depleted and the system reached a completely unsustainable state. The fraction of sustainable agents of a run n_s and the average group attitude A are constructed in the same way. Here, $n_s = 0$ or $A = 0$ would indicate that all individual or groups have an unsustainable group attitude, while $n_s = 1$ or $A = 1$ would indicate that they all carry a sustainable attitude. Then, for each of the main observables, a macroscopic ensemble average is constructed over all runs $j = 1, 2, \ldots, n_r$:

$$\langle n_s(t) \rangle = \frac{1}{n_r} \sum_{j=1}^{n_r} n_{s,j}(t). \tag{11}$$

Therefore, $\langle n_s \rangle$ denotes the average fraction of sustainable individuals for a parameter set and its runs. Correspondingly, the macroscopic ensemble average for the stock reads as

$$\langle S(t) \rangle = \frac{1}{n_r} \sum_{j=1}^{n_r} S_j(t). \tag{12}$$

In the following, the average stock $\langle S \rangle$ and the average fraction of sustainable individuals $\langle n_s \rangle$ will be the main quantities of interest. Their state at the end of a simulation determines the average outcome of the system for a certain parameter set. To distinguish averages in one run and ensemble averages over all runs, S and n_s will be called micro quantities, while $\langle S \rangle$ and $\langle n_s \rangle$ will be named macro quantities. For plotting, the quantities will be shown in %.

3.1 Groups & group memberships

First, the influence of the number of groups N_g and the number of memberships per agent N_m on the full model is analysed. Both parameters are varied while the others are kept fixed at values for which the forcing to fall into a sustainable or unsustainable state is balanced, $\theta_{DN} = 0.5$, $\theta_{IN} = 0.5$, $\Delta T_a = 1$, $\Delta T_g = 1$, such that the focus lies on N_g and N_m alone. For all combinations of N_g and N_m, the outcome varies around $\langle n_s \rangle \approx 0.5$ with no qualitative differences for different parameter combinations. Additionally, a very high standard deviation is observed. This small parameter sensitivity for N_g and N_m can be explained by the fact that individuals and groups are initialised as sustainable or unsustainable with the same probability of 0.5, such that thresholds of $\theta_{DN} = 0.5$ and $\theta_{IN} = 0.5$ quickly lead a simulation to fall either into a sustainable or unsustainable state, as both norm processes eventually align themselves. This splitting up into two extreme states for multiple runs also explains the high standard deviation and can be seen in Fig. 4, columns 1 and 3. No process of the norm framework is dominant due to the balanced weights, such that the initial distribution causes the outcome, depending on which norm uptake threshold θ_{DN} or θ_{IN} it crosses to which degree. Thus, N_g and N_m do not have a qualitative influence on the average sustainability outcome of the model on an ensemble average level. Still, on the micro level, interesting dynamics arise that depend on N_g and N_m. These dynamics are relevant to real-world situations as detailed below, since there is not an ensemble of realities the agents live in and aim to manage, but one system trajectory where it greatly matters whether a complete collapse of resources occurs or not.

3.1.1 Analysis of micro dynamics

To analyse these micro dynamics, two combinations of group sizes and group memberships per agent, that represent idealised cases, will be investigated more thoroughly.

The first combination of which the micro dynamics are investigated consists of two groups $N_g = 2$ and $N_m = 1$; i.e. each agent is member in one group only. Here, groups have an average degree, i.e. average number of members, of 200. This case represents a system in which two large groups exist, that do not overlap in their member base. An example in which individuals choose between two groups (or ideologies, etc.) is e.g. a two-party political system.

The second combination is $N_g = 64$ and $N_m = 4$, where each group has an average number of members of 25. This case represents a system with many smaller groups that can overlap. This represents a case with a high number of smaller groups, such as (workers) associations, clubs or similar, that take a stance on a relevant behaviour.

Figure 4 shows the time series for $t = 500$ time steps of the average quantities of n_r runs for $N_g = 2$ with $N_m = 1$ and $N_g = 64$ with $N_m = 4$ for the standard case and the special case that the group attitudes A do not change in time, such that groups always send the same injunctive norm, i.e. becoming a vessel for fixed norms. After $t = 500$ all trajectories that can reach a dynamic equilibrium have converged to it.

It can be seen that in the case with dynamics group attitudes (Fig. 4 columns 1 & 3), the average fraction of sustainable agents $\langle n_s \rangle$ quickly splits up into two levels between multiple runs, at approximately 0.25 and 0.75. During a very dynamic beginning phase, both norms eventually align, leaving a single run to end up in a sustainable or unsustainable state. The found probabilities agree with the probabilities expected by Eq. (9).

A network of agents that has overlapping group memberships ends up in one of the two states more quickly (column 3) than a system with two groups only that do not share any members and where the connection is only given on one of the two layers (column 1). This can be explained by the fact that for $N_m = 1$, the two member bases are not connected. Therefore, there is no communication between the two groups and both systems continue to co-exist in a "polarised" state, such that one group can temporarily continue sending a norm that might be differing from the majority behaviour, slowing down the convergence to one majority behaviour.

3.1.2 Fixed group attitude

Interesting dynamics arise when the group attitude is fixed (i.e. A is not allowed to change), which mimics a fix injunctive norm or an injunctive norm entity in the socio-cultural domain. When looking at the micro behaviour of $N_g = 2$ with $N_m = 1$ (Fig. 4 column 2) and $N_g = 64$ with $N_m = 4$ (Fig. 4 column 4) for such a fixed group attitude, it can be seen that instead of having a split up into only two distinct levels, intermediate states can be found for the average behaviours n_s.

For $N_g = 2$, two extreme states are still reached where both groups are initialised with the same group attitude, such that both their norms "point" into the same direction, forcing the descriptive norm to align. The bifurcation into intermediate levels is found in runs where one group has $A = 0$ and one has $A = 1$, such that both injunctive norms can prevail. The intermediate state then depends on the prevalent descriptive norm and the individual learning, which is encoded by the continuous harvest h, thus reaching continuous levels of n_s.

For $N_g = 64$ with $N_m = 4$ (Fig. 4 column 4), $\langle n_s \rangle$ centres around a mean expectation value $\mu = \frac{N_g}{2}$ as it can be described by a Binomial distribution: there are N_g groups with either state $\{0, 1\}$. The possible combinations that can be reached are given by a combination with repetition, where the order does not matter. As the attitudes of groups are initialised as a Bernoulli process, the probability to obtain a certain number of sustainable groups in a run is approximately given by a Binomial distribution.
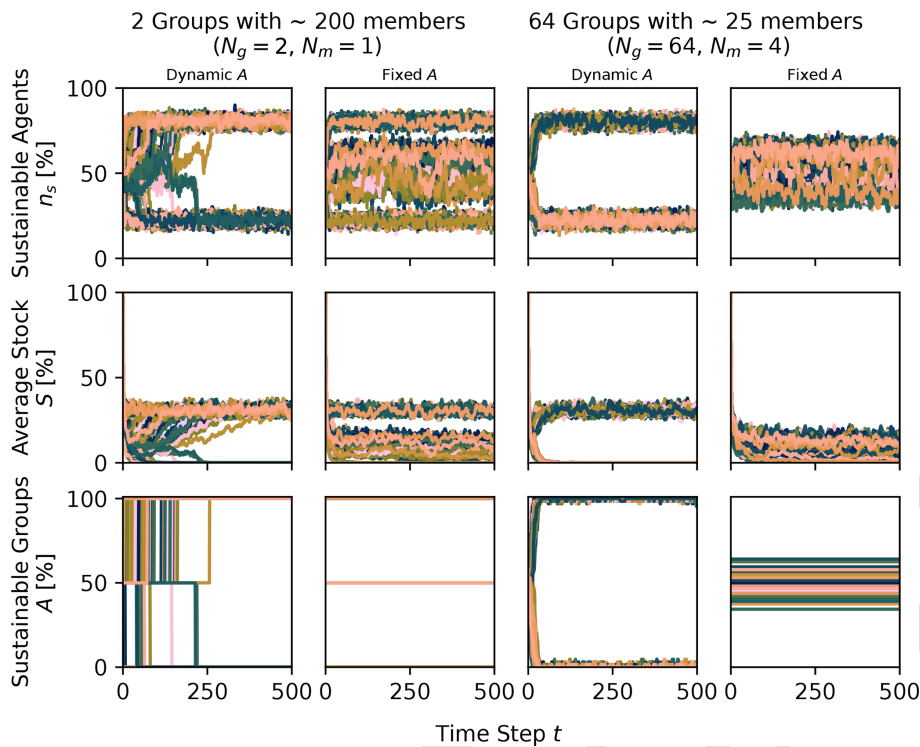
Figure 4. The time series for 500 time steps of the average fraction of sustainable individuals, $\langle n_s \rangle$ (row 1), the average stock, $\langle S \rangle$ (row 2) and the average group attitude A (row 3). Plotted are 100 runs, colour-coded (using scientific colour maps as described in Crameri et al. (2020), in this and in following Figs.), for different group constellations (numbers of groups N_g, number of memberships in different groups of an agent N_m and the resulting approximate group size; columns 1–2 vs. 3–4) and the special cases where A does not change in time, i.e. fixed "injunctive norms".

This example showcases how strong (static) social groups can structure the behaviour space into different levels. However, in general, groups can change their behaviour dynamically, such that we focus on dynamic group attitude for the analysis of the thresholds and the time scales in the model.

3.2 Thresholds

We now analyse the influence of the uptake threshold for the descriptive norm θ_{DN} and the threshold for group attitude change, which represents the threshold for the injunctive norm, θ_{IN} on the behaviour of the model. Figure 5 shows the average fraction of sustainable agents $\langle n_s \rangle$ and the average stock $\langle S \rangle$ for a variation of θ_{DN} and θ_{IN} in the group constellations of $N_g = 2$ with $N_m = 1$ and $N_g = 64$ with $N_m = 4$, while $\Delta T_a = 1$, $\Delta T_g = 1$. The result is shown after $t = 100$ time steps, after which all runs have reached a dynamic equilibrium.

As expected, a regime shift arises along a diagonal axis, where θ_{DN} and $\theta_{IN} \approx 0.5$, which divides the state space into two areas, one sustainable and one unsustainable. Along this axis, the system behaves as observed before and is very susceptible to the initial distribution of states in the agents. Where θ_{DN} and θ_{IN} are small, the system expectedly con-

verges into a sustainable state, while the opposite is true for θ_{DN} and θ_{IN} are large. Interesting is that the transient regime extends beyond the diagonal axis, splitting up towards the upper left and lower right corners of the panel. This transient regime occurs when one of the two thresholds is large and the other small. When analysing micro runs, it can be seen that in these cases, both norm processes approximately cancel each other out, for example when a strong sustainable injunctive norm and a strong unsustainable descriptive norm both prevail and keep influencing agents to change their behaviour with similar strength. This leads the system to stochastically move around $\langle n_s \rangle \approx 0.5$. The corresponding stocks in Fig. 5 exhibit low, but non-zero values. This is found for all parameter combinations that lead the agents to spend comparable time with the high and low harvesting effort levels, due to the fact that recovery of the resource in this model is slower than exploitation. Therefore, three main regimes are introduced by the two thresholds in the social norm model: a sustainable regime, an unsustainable regime, and a transient regime. They are divided by regime shifts with increased values of $\sigma_{\langle n_s \rangle}$. Due to the low stock of the transient regime, only the completely sustainable regime can be considered a safe operating space.

For $N_g = 2$ with $N_m = 1$, the same qualitative behaviour as for $N_g = 64$ with $N_m = 4$ is found, but the transition is smeared out in the direction of θ_{IN}: since $N_m = 1$, all agents are member of one group. Thus, this one group can exert a strong influence on all of its members until it undergoes the first group update. The probability to "escape" the threshold scales with θ_{IN} and leads to a low gradient around $\theta_{IN} \approx 0.5$.

If the number of groups is increased, a single group might still be able to "escape", but will not influence the outcome of the ensemble average as much, as its relative importance is only a fraction of the number of groups. This explains why the transition is more abrupt for $N_g = 64$ with $N_m = 4$.

The results show that the major structuring parameters for the resulting safe operating space of the model are the two thresholds of the social norm framework. Different group constellations also influence the shape of the transient regime. Therefore, the social norm and group constellations structures the sustainability space and the safe operating spaces in the model.

3.3 Timescales

Now, the influence of the parameters that influence the timescales on which individual and group updates take place are compared for their influence relative to each other and the total simulation time. Note that those properties are dimensionless here and scale only with each other. Figure 6 shows the behaviour of the average fraction of sustainable individuals $\langle n_s \rangle$ for both group constellations. The result is shown after $t = 100$ time steps, after which all runs have reached a dynamic equilibrium for the analysed parameter ranges.

For a number of groups $N_g = 64$, where each agent is member in $N_m = 4$ groups (Fig. 6a) and small values of the individual updating timescale ΔT_a ($\Delta T_a < 1$), the model tends to become more unsustainable, $\langle n_s \rangle \approx 0.2 - 0.4$, than the expected $\langle n_s \rangle \approx 0.5$ for thresholds $\theta_{DN} = 0.5$ and $\theta_{IN} = 0.5$. This finding is in line with the results of the original Exploit model (Wiedermann et al., 2015) and its extension (Barfuss et al., 2017), in which myopic agents also increased the probability to end up in an unsustainable state. With increasing values of ΔT_a relative to ΔT_g, the system first settles down in an equilibrated outcome (due to a bifurcation into two states again) with $\langle n_s \rangle \approx 0.5$ and then actually exhibits a slight tendency towards sustainable outcomes.

This can be explained by the added influence of the own harvesting success through the individual learning: an enhanced ΔT_a increases the rate at which agents are influenced by their own harvest in taking a decision. We computationally and analytically find that decision-making influenced by the individual learning of harvest alone, without social norms, leads to a state of $n_s = 0.5$ and $S = 0$ (see Appendix A). The nudge towards an unsustainable outcome due to the harvesting influence is enough to tip the social norm into an unsustainable direction as well in the dynamic beginning phase of a run. As soon as both norm processes point

in the unsustainable direction, the system will not recover. This is in agreement also with the behaviour of the harvesting model without the normative influences, where fast social dynamics lead to a system that ends up in an unsustainable fixed point after a critical slowdown. Even if this influence makes up a fraction of only 0.33 of the decision-making argument, against two processes that are not qualitatively time-dependent, it adds up to the observed behaviour.

In the other direction ($\Delta T_a > 2$), the fact that the system actually has a slight tendency to be sustainable can be explained by the low update rates and the decreased, but still existing, influence of utility-maximising individual learning. Because of that, the system quickly approaches a state in which $\langle n_s \rangle \approx 0.5$ and $\langle S \rangle \approx 0.25$ in the beginning of a run, as on average no update has taken place yet. When the first update takes place, unsustainable individuals will completely have exploited their stock, increasing their probability to become sustainable, while sustainable individuals will have more stock, making them less likely to switch behaviours as compared to their unsustainable counterparts. Therefore, the sustainable strategy is slightly favoured, which leads to the tendency towards a sustainable outcome.

For $N_g = 2$ with $N_m = 1$ (Fig. 6b), the transition is only found at smaller values of ΔT_a. This can be explained by the micro dynamics for said constellation that was found before to exhibit a much slower convergence towards two extreme states, thus limiting the influence of the harvest that it most detrimental in systems with quick social dynamics.

Therefore, ΔT_a is found to introduce a regime shift into the system, even though it is not as pronounced, in the sense that it does not divide between a completely unsustainable and a completely sustainable regime, but rather between a more unsustainable and a more sustainable one.

It is also found that the transition regime size diminishes in ΔT_a-direction for increasing the timescale of group updates ΔT_g. Groups update more slowly and they are able to keep their attitude and their influence on member individuals for a longer time (approximating the fixed A-case), reducing the exploitative effect of the individual learning model. Very high values of ΔT_a and ΔT_g compared to t are omitted here, since the system does not converge, due to the resulting large updating times that allow no dynamics to arise during t.

For constellation $N_g = 2$ with $N_m = 1$, the effect is qualitatively similar but reduced. This can be explained by the longer convergence time for single groups for $N_g = 2$ with $N_m = 1$.

3.4 Group attitude intervention

We now mimic a real-life situation in which a non-specified policy intervention leads some groups to change their group attitude and thus send a different injunctive norm to their members. This scenario showcases the possibilities of research questions that can be addressed with Nexploit and related models.
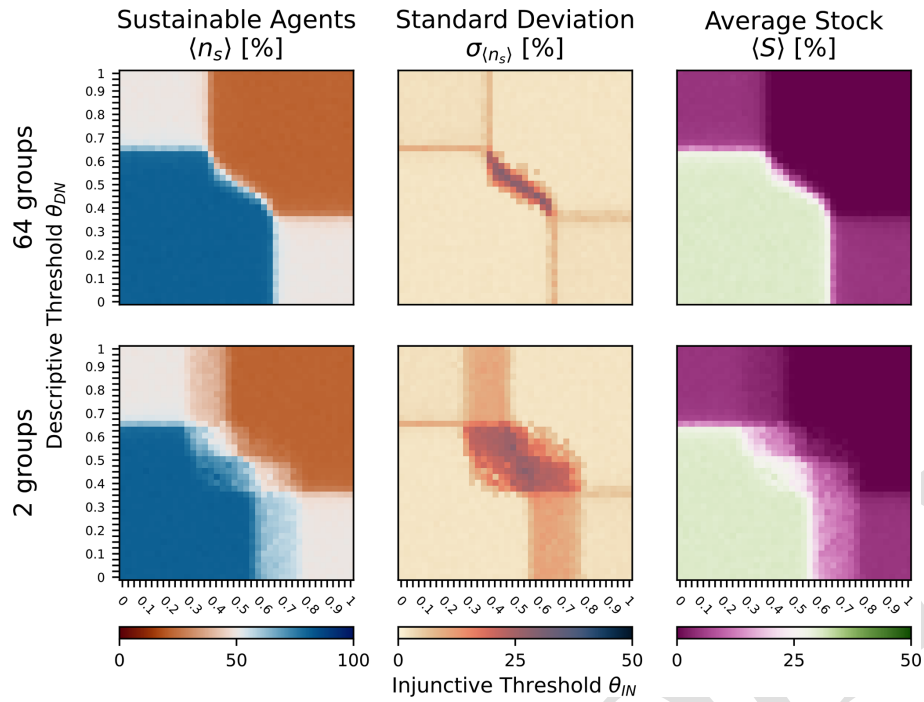
Figure 5. The influence of the uptake thresholds for the descriptive and injunctive norm, θ_{DN} and θ_{IN}. Top: the ensemble average fraction of sustainable individuals for 100 runs, its standard deviation and the resulting stock levels, for $N_g = 2$ groups, where an agent is member of $N_m = 1$ groups each. Bottom: the ensemble average fraction of sustainable individuals, its standard deviation and the resulting stock levels, for $N_g = 64$ with $N_m = 4$.
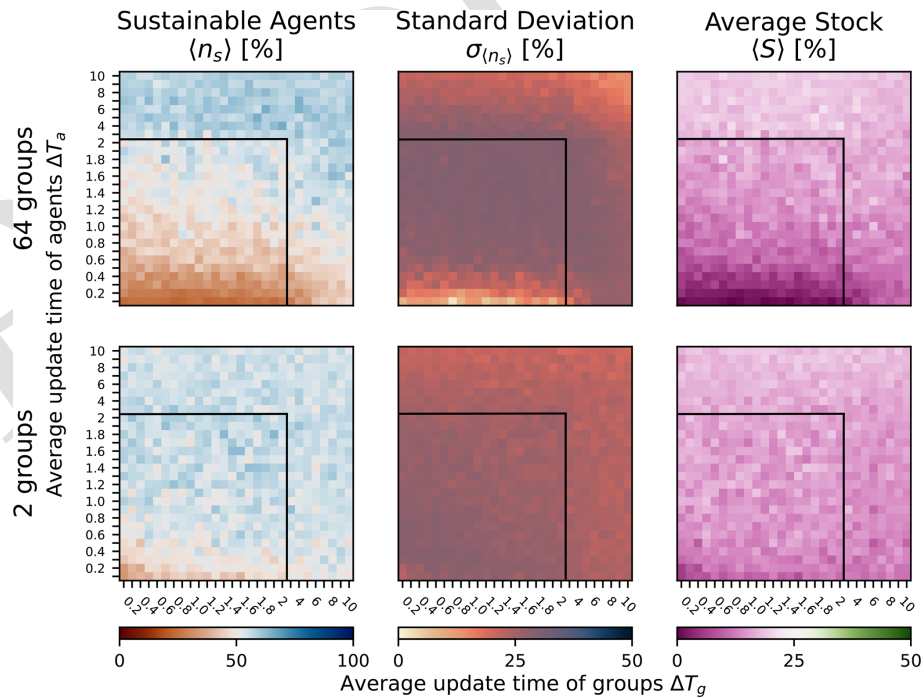
Figure 6. The influence of the updating timescales of the individuals and social groups, ΔT_a and ΔT_g, after 100 time steps. Top: the ensemble average fraction of sustainable individuals for 100 runs, its standard deviation and the resulting stock levels, for $N_g = 2$ groups, where an agent is a member of $N_m = 1$ groups each. Bottom: the ensemble average fraction of sustainable individuals, its standard deviation and the resulting stock levels, for $N_g = 64$ with $N_m = 4$. Please note the change in resolution.
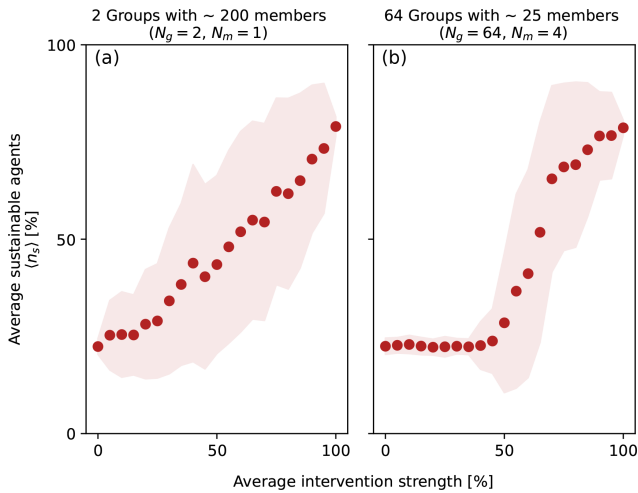
Figure 7. The ensemble average fraction of sustainable individuals $\langle n_s \rangle$ over 100 runs after 100 time steps against different values of the fraction of policy-influenced groups (indicating the strength of an arbitrary policy intervention) for **(a)** $N_g = 2$ groups, where an agent is member of $N_m = 1$ groups each, and **(b)** $N_g = 64$ with $N_m = 4$. The shaded area around the data points indicates the respective standard deviation $\sigma_{\langle n_s \rangle}$.

In the beginning of a simulation run, the system is set to an unsustainable state, that is, all individuals and all groups have an unsustainable attitude and the descriptive and injunctive norms point into the unsustainable direction. At $t = 50$ of a run, resources $\langle S \rangle$ will almost be depleted. Then, a changing fraction of groups will switch their group attitude to sustainable and keep it for $dt = 10$ until they are allowed to change their attitude again, to ensure that the intervention can have an effect. The fraction of policy-influenced groups is determined by a probability for each group to switch, yielding a measure of the intervention strength when averaging over all group and all runs. It is then observed until $t = 100$ (after which the model has converged to a basin that it does not leave any more) whether the system falls back into an unsustainable state or whether the intervention can carry it into a sustainable state. The choice of the other parameters is $\Phi_a = 0.25$, $\Delta T_a = 1$, $\theta_{DN} = 0.5$, and $\theta_{IN} = 0.5$.

This setting can be thought to represent the current pathway of many real-life systems, in which ecological resource dynamics and socio-cultural dynamics are coupled: such systems are often over-exploited. Then, for example, because of the realisation of an imminent collapse, groups themselves might change their attitude towards exploitation. Or social institutions that are not included in this model, such as governments or political actors, might encourage groups to send a different normative message towards their members to encourage them to change their behaviour to allow the overall system to transition into a more sustainable mode of operation.

Figure 7 shows the resulting average fraction of sustainable agents for 100 runs, depending on the intervention strength and the group constellations. In the left panel, with $N_g = 2$ groups, where each agent is a member of $N_m = 1$ group, the probability to find the agents in a sustainable state at the end of a run increases in an almost linear fashion with the intervention strength. This shows the expected result that targeting more groups is more likely to induce social change than targeting few groups.

In the left panel, for intervention strengths < 0.5, all runs end up in an unsustainable state. The number of groups that change their state is thus not enough to create a lasting effect. For an intervention strength > 0.5, the probability to find the system in a sustainable state at the end of a run increases in a non-linear fashion, resembling a regime shift.

The difference in the two systems can be explained by the fact that for two large, non-connected groups, as for $N_g = 2$ with $N_m = 1$, with linearly increasing probability, the probability for the groups to switch increases linearly as well. When one group switches its state because of the intervention, it has enough time to influence its members via the injunctive norm, such that the system does not switch back. It directly adds to the resulting ensemble average. For many connected groups, as for $N_g = 64$ with $N_m = 4$, groups that become sustainable also influence their members. But as members might still be connected to unsustainable groups, the number of groups that change attitude must be large enough to have an effect on the overall system. This introduces the found regime shift in the system with $N_g = 64$ with $N_m = 4$, which depends θ_{DN}. When enough group members change their behaviour during the intervention time to surpass the descriptive norm threshold, the whole system will become sustainable.

This brief analysis shows that, in our model, systems react completely differently to a stylised intervention depending on the group structure. This highlights the importance of quantitatively modelling this behaviour and understanding the dynamics, rather than assuming that attitude changes will scale like-ways in different group structures.

3.5 Discussion

The micro dynamics of the model were found to exhibit trajectories that split up into two extreme states. This is a common effect in coupled social–ecological models with majority-enforcing social norms, where typically extreme equilibria of a single remaining behaviour are found (Farahbakhsh et al., 2022). This behaviour implies an extreme outcome such that it has the potential to either support a sustainable outcome or lead the system to completely collapse.

In our study, we, however, observe a third regime occurring when both norm processes approximately cancel each other out. Therefore, the modelling of two distinct but interacting dynamic norm processes actually introduces intermediate equilibria. This shows that consideration of both de-

scriptive and injunctive norms, beyond common parametric representations (Farahbakhsh et al., 2022), is vital in understanding the complex dynamics of social norms.

The main process governing the outcome is the threshold behaviour of norm adoption. We found that only by reducing the thresholds of both norm aspects one arrives at a fully sustainable state, as the equilibrated state does not have sufficient stock left to be considered sustainable. A dynamic and detailed representation of norms therefore is crucial when modelling policy norm interventions, which are said to be highly contextual (Constantino et al., 2022).

The regime shifts in the state space due to these thresholds could be thought of as representing a social tipping point. While a simple threshold in a model of course does not include all criteria for a social tipping point, nor represents its complexity (Winkelmann et al., 2022), it presents a good starting point for the investigation of tipping in social norms.

The inclusion of groups and different group membership structures introduces more diverse outcomes into the model. Typically, systems with high numbers of groups N_g and group memberships N_m were found to exhibit stronger regime shifts, while systems with low numbers of groups and group memberships exhibit less sharp transitions. Considering regime shifts as proxy for social tipping points, tipping a two-party political system in the model would require a stronger effort in altering the social norm defining parameters (e.g. the uptake thresholds for sustainable norms) than in a more fractured, multi-group system. Having persistent (injunctive) norms mediated by groups, represented in the model by not allowing groups to update their attitude (i.e. very change-inert groups), introduced multiple distinct levels of stable states that trajectories approached. Social norm systems are said to typically have such multiple equilibria (Young, 2015). Considering groups in social norm contexts thus plays an important role when trying to capture the full social dynamics, as suggested in the literature (Constantino et al., 2022).

Looking at the updating timescales, the system shows a regime shift between a rather sustainable and a rather unsustainable regime, where high updating rates point towards a more unsustainable outcome because of the increased influence of utility-maximising individual learning. Thus, the more often agents consider their own economic utility, the less likely the system is to end up in a sustainable state. Short-term decision-making that is based on current success leads to augmented utility-maximising behaviour and unsustainable outcomes. A less myopic approach lets agents observe the advantage of a sustainable strategy, increasing the probability of a sustainable outcome, a finding that is in line with similar analyses (Wiedermann et al., 2015; Barfuss et al., 2017; Geier et al., 2019).

We found that socially inert (increased ΔT_g) groups can dampen unsustainable tendencies. For high updating rates instead, groups tend to simply follow an unsustainable trend. Making groups resistant to such trends, might increase the resilience of the system against unsustainable tendencies because of egoistic considerations, as well as help overcome detrimental utility-maximising behaviour or social inertia.

A system with two non-connected large groups, combination $N_g = 2$ with $N_m = 1$, takes longer to reach an extreme state than a system with many connected small groups, combination $N_g = 64$ with $N_m = 4$. This might provide one with more time to counter unsustainable tendencies in the case of $N_g = 2$ with $N_m = 1$, but on the other hand, might prevent the system from falling into an unsustainable state fast enough when under time pressure and vice versa for $N_g = 64$ with $N_m = 4$. Also, in real-life systems, groups, social norms, and individual learning crucially influence the time scales that determine the success in mitigating anthropogenic climate change (Otto et al., 2020).

We simulated an intervention to provide an idea of the capabilities of the model. For two choices of N_g and N_m, i.e. two distinct group network topologies, the system responds differently depending on the intervention strengths. In the case of $N_g = 2$ with $N_m = 1$, which could represent a polarised two-party system, the outcome followed the intervention strength in a linear fashion. Of course, changing the attitudes of both large groups yields the largest success. But this also shows that changing the influence of one group on its members can already bring about positive change. Still, in a real-life system, this may be more costly than influencing smaller groups. For $N_g = 64$ with $N_m = 4$, a non-linear response was found, exhibiting a regime shift. This suggests that in connected systems with many small groups, interventions must reach a sufficient large number of groups in order to overcome unsustainable descriptive norms and detrimental utility-maximising behaviours or social inertia. Otherwise, it might simply be a waste of resources. This points to the fact that idiosyncratic features of a system need to be considered, i.e. in policy-design (Constantino et al., 2022).

Finally, with the model results, we confirmed the assumption that social norms can be crucial to the overall dynamics in resource models (Satake et al., 2007; Lade et al., 2013; Sigdel et al., 2017; Farahbakhsh et al., 2022). Even in a simplistic social–ecological model, the detailed social norm framework reproduces important, basic characteristics of social systems under normative influence, while having additional possibilities of shaping and structuring safe operating spaces of resource management.

4   Conclusions

In this work, a coupled social–ecological model simulating resource use, taking into account social norms and groups in a social norm framework, was designed, developed and tested in the copan:CORE framework. It consists of a stylised ecological resource model coupled with a social norm component in a multi-layer network. The entity type "Group" was introduced into the copan:CORE framework with this work

as well. The new social norm framework, consisting of a descriptive and an injunctive norm mechanism, was included as a component to copan:CORE and has been tested. A mechanism to model individual learning via harvesting success was presented and analysed.

We have found that group parameters can structure the solution space into various levels of sustainability, representing a more realistic set of equilibria than mere "black or white" collapse or sustainability scenarios. The adequate consideration of group topologies and norm structures can be crucial for the outcome of social norm interventions in the model.

We have shown that this multi-layer network framework for modelling social norms and groups allows for detailed investigation of the influence of direct descriptive and, in particular, injunctive normative and group-related social processes in feedback with ecological processes. Taking extended normative and group-related social processes in feedback with ecological processes into account can crucially alter the dynamics of social–ecological models and should thus be considered on the path to developing more detailed representations of socio-cultural systems in integrated world–Earth modelling. Reaching a sustainable state, i.e. a safe operating space, in the coupled model becomes more likely with low thresholds for conforming to sustainable norms, as well as lower consideration rates of own resource harvesting success. Although still highly stylised, the social norm framework substantially extends modelling possibilities, as compared to one-dimensional parametric representations.

Limitations & outlook

In this work, we have focused on certain parameters and their combinations and ranges that exhibited interesting dynamics. This is possible due to their dimensionless nature in this model. For future model fits to real world data it will be necessary to do a more thorough sensitivity analysis.

While models of social norms from the computational social sciences might exhibit greater levels of complexity in the representation of single norm-relevant processes (Andrighetto and Vriens, 2022), our model of endogenous descriptive and injunctive norms clearly fills a gap in the modelling of norms in a social–ecological systems context.

In future work, it is desirable to include additional aspects that have already commonly been modelled in other agent-based models. These aspects are: (i) norm conformation (McDonald and Crandall, 2015) or internalisation (Henrich and Ensminger, 2014; Elsenbroich and Gilbert, 2015; Gavrilets and Richerson, 2017; Gavrilets, 2020) and (ii) norm enforcement or sanctioning (Opp, 2001; Tavoni et al., 2012; Elsenbroich and Gilbert, 2015; Nhim et al., 2019). So far these aspects have been left out or have only been aggregated into probabilistic effects via the decision-making function; i.e. it is assumed that there must be some social pressure for agents to adhere to social norms, be it social or non-social sanctions. Such aspects could be included next.

For example, sanctioning could be introduced into the model, by reducing the harvest of an agent that does not follow the social norm, similar to the taxation mechanism in Geier et al. (2019), possibly introducing a new leverage point for introducing a prevalent sustainable behaviour in the system. A sanctioning mechanism could be interesting in particular for a common-pool resource setting, which is not analysed here in favour of a personal pool setting. This choice constrained any additional information flow to the socio-cultural level, as individuals only retrieve information about the harvesting behaviour of their neighbours via the norm mechanism, not via the state of a common pool resource. In this study this was a desired feature, so as to set the focus on the social norms, but in future studies we will apply a common-pool setting as well. Additionally, the private-pool setting kept the results comparable to former works whose setup for the resource growth and harvesting dynamics we built upon.

It is important to note that separating descriptive and injunctive influences is not always trivial and there is an overlap between them (Bicchieri, 2017). Therefore, the interaction on both network layers that we treat as distinct descriptive and injunctive norms might contain aspects of the other type in reality. Additionally, both networks might be related, such that groups of work colleagues might coincide with friend circles and so on. There might also be other important in- and out-group effects that we miss due to this overlap (Restrepo-Plaza and Fatas, 2022). Introducing such effects might further reduce dichotomous outcomes.

So far, the behaviour of the groups' members is only translated into a normative attitude by the groups through a process that we aggregate by a threshold. In fact, literature on attitude–behaviour gaps suggests that it might often be the case that behaviour and attitude do not translate into one another, for example in ethical consumption (Carrington et al., 2010). As a next step, one might consider extending the framework by giving agents an attitude as another feature, which then relates to the injunctive norm. This could further improve the representation of the injunctive norm, as it could disentangle the direct influences on behaviour, putting a clearer focus on the different foci and flavours of descriptive and injunctive norms.

As an additional next step, the network structures might be remodelled in two ways: first, social structures could be more realistically represented by the current random networks, for example including homophily (McPherson et al., 2001) or preferential attachment (Barabási and Albert, 1999). Second, the network structures could be adaptive, enabling changing links between agents and agents or groups and agents (Gross and Blasius, 2008). This would also allow for the modelling of dynamic group formation, which plays an important role in the emergence of normative behaviour (Ehrlich and Levin, 2005) and could also be used to include ostracism as way of including sanctioning (Elsenbroich and Gilbert, 2015). It would also allow for an apt inclusion of group membership

effects, for example, people might be more likely to follow the behaviour of an in-group neighbour (Spears, 2021).

We leave such considerations to future work with an adaptive network, which was wilfully left out here as not to overshadow the dynamics of the norms.

In this study, we have successfully included a nuanced social norms framework with groups into an existing social–ecological network model of natural resource use and demonstrated the capabilities and relevance of the framework. Technical advances and insights from this work can be used in the development of more process-detailed world–Earth models (Donges et al., 2020), for example for studying the biophysical potentials and social spreading opportunities for regenerative agriculture (Breier et al., 2023). The social norm framework can now be used to study diverse problems in the context of social norms, such as energy consumption, where descriptive and injunctive norms play an important role in the promotion of energy conservation (Bonan et al., 2020a). When combined with further components of the copan:CORE framework, it can contribute to developing more nuanced world–Earth models, pursuing vaster possibilities to study global challenges and design policy interventions based on this knowledge (Steffen et al., 2020; Donges et al., 2020).

Appendix A

To understand the influence of the sole individual learning via harvesting dynamics, one can analyse its macro quantities analytically and conduct a stability analysis, which is possible under some simplifying assumptions. Here, the probability to switch behaviour in a social update depends only on the harvesting success of the individual, mimicking utility-maximising behaviour or social inertia. Setting w_{IN} and w_{DN} to 0, while $w_h = 1$, yields a probability to switch of

$$P\left(E_j \to E_k\right) = \frac{1}{1 + \exp(-k \cdot -h')}. \tag{A1}$$

This means that agents will be less likely to switch their behaviour when they are currently successful, in line with the notion of a utility-maximising win–stay, loose-shift strategy.

A1 Derivation of dynamical system

The relevant dynamics on the agent level can be summarised by the average stock in the system S and the average fraction of sustainable agents n_s. The goal is to find the equations of motion for both quantities, such that a two-dimensional dynamical system with an analytical solution is found. The average fraction of sustainable agents n_s and the average fraction of unsustainable agents n_u are connected by

$$n_s = 1 - n_u. \tag{A2}$$

First, the dynamics of the social process are assumed to be fast, that is instantaneous, in comparison to the natural dynamics. Thus, in an infinitesimal time step $(t, t + dt)$, the expected fraction of sustainable agents is given by

$$dn_s(t) = dn_{us}(t) - dn_{su}(t), \tag{A3}$$

where dn_{ns} and dn_{sn} represent the fractions of agents that switch their behaviour and, correspondingly, effort, from unsustainable to sustainable and vice versa. The explicit notion of time dependence is dropped from now on.

If all agents are assumed to switch instantaneously, the changing fractions dn_{us} and dn_{su} are given by the fraction of agents that are found in one state (e.g. n_u) times the probability to leave said state since this probability is the same for all agents embodying one behaviour. Hence,

$$dn_{us} = n_u \cdot p_{u\to s}, \tag{A4}$$

where $p_{u\to s}$ denotes the probability to switch from an unsustainable to a sustainable state. The same holds in the other direction and

$$dn_{su} = n_s \cdot p_{s\to u}. \tag{A5}$$

Putting Eqs. (A5) and (A4) into Eq. (A3), one obtains

$$dn_s = (n_u p_{u\to s} - n_s p_{s\to u})dt. \tag{A6}$$

When inserted with the mapped sustainable and unsustainable harvest of

$$h'_u = \frac{2}{E_u} \cdot E_u S_u - 1 \quad \text{and} \quad h'_s = \frac{2}{E_u} \cdot E_s S_s - 1, \tag{A7}$$

the probabilities to switch efforts are

$$p_{u\to s} = \frac{1}{1 + \exp(k(2S - 1))} \quad \text{and}$$

$$p_{s\to u} = \frac{1}{1 + \exp(k(\frac{2E_s S}{E_u} - 1))}, \tag{A8}$$

where the system is further simplified by assuming the agents to harvest from one common average stock $S = S_u = S_s$ of the system, instead of their own resource stock. For this, it is presumed that the social dynamics are fast enough to equilibrate the differences in the average stock of sustainable individuals and unsustainable agents. This notion also ignores any fraction of stock that switching agents might bring into the other pool. In an infinitesimal time interval $(t, t + dt)$, an average harvest is extracted from the average stock in the system, according to Eq. (3). This harvest term depends on the average fraction of agents that adopt each harvesting behaviour:

$$dS = (S(1 - S) - S(n_u E_u + n_s E_s))dt. \tag{A9}$$

Applying Eq. (A2), one can find the equations that govern the dynamic system:

$$\frac{dS}{dt} = S(1 - S - E_u - n_s E_u + n_s E_s), \tag{A10}$$

$$\frac{dn_s}{dt} = (1 - n_s)p_{u \to s} - n_s p_{s \to u}. \tag{A11}$$

Note that Eq. (A11) includes its dependency on the stock through the probabilities $p_{u \to s}$ and $p_{s \to u}$.

A1.1 Fixed points & stability

The set of two coupled differential Eqs. (A10) and (A11) fully describe the approximate, macroscopic time evolution of the harvest model alone. They form a two-dimensional dynamical system:

$$\dot{\boldsymbol{u}} = \boldsymbol{f}(\boldsymbol{u}), \qquad \boldsymbol{u} = (u_1, u_2). \tag{A12}$$

In this case, $u_1 = S$ and $u_2 = n_s$. Such dynamical systems are better understood through stability analyses, a tool that will be applied in the following. The analysis follows Roth-Fauchere (2023), including only the most important steps of the process without further theoretical introduction.

When working with dynamical systems, one typically searches for fixed points, where the time evolution vanishes and the system state stays constant in time: TS4

$$\dot{\boldsymbol{u}}_0 = \boldsymbol{f}(\boldsymbol{u}_0) = 0. \tag{A13}$$

In two-dimensional systems, one finds nullclines

$$f_i(\boldsymbol{u}) = 0, \qquad \dot{u}_i = 0, \tag{A14}$$

lines where one of the equations vanishes. The location in which nullclines intersect then gives the fixed points of the system. Setting

$$\frac{dS}{dt} = 0 \quad \text{and} \quad \frac{dn_s}{dt} = 0,$$

the nullcline of n_s is found to be

$$n_{s,0} = \frac{p_{u \to s}}{p_{u \to s} + p_{s \to u}}. \tag{A15}$$

S is found to have two nullclines:

$$S_0^1 = 0 \quad \text{and} \quad S_0^2 = 1 - E_u + n_s(E_u - E_s). \tag{A16}$$

Note that the superscript denotes the number of the nullcline, not an exponent. The fixed point is then given by the intersections of the nullclines, $n_{s,0} \cap S_0^1$ and $n_{s,0} \cap S_0^2$, respectively. The first intersection is found by inserting Eq. (A16) into Eq. (A15):

$$n_{s,0}(0) = \frac{\frac{1}{1+e^0}}{\frac{1}{1+e^0} + \frac{1}{1+e^0}} = \frac{1}{2}, \tag{A17}$$

yielding the first possible fixed point

$$\boldsymbol{u}_0^1 = (0, 0.5). \tag{A18}$$

To find the intersection $n_{s,0} \cap S_0^2$,

$$n_{s,0} = S_0^2 \tag{A19}$$

is solved for S, which then can be used to find the corresponding value of n_s. The analytical solution of Eq. (A19) is not trivial. Hence, the Taylor expansion of $n_{s,0}$ is used to simplify the equation, which then reads as

$$n_{s,0} \cong \frac{1}{2} + S\ell + \mathcal{O}(S^2), \quad \ell = \frac{2k(E_s + E_s e^k - e^k - 1)}{E_u(2 + 2e^k)^2}. \tag{A20}$$

Only the terms up to first order are used and terms of higher order are omitted. Solving Eq. (A19) leads to the second possible fixed point

$$\boldsymbol{u}_0^2 = (x, n_{s,0}(x)), \qquad x = \frac{1 - E_u - \frac{E_s - E_u}{2}}{\ell(E_s - E_u) + 1}. \tag{A21}$$

For the parameter choice of $E_u = 1.5$ and $E_s = 0.5$ in this model and $k = 3$, the second fixed point coincides with the first one:

$$\boldsymbol{u}_0^1 = \boldsymbol{u}_0^2 = \boldsymbol{u}_0 = (0, 0.5). \tag{A22}$$

Of prime interest when dealing with fixed points is their stability, that is the development of trajectories that start close to the fixed point at $\boldsymbol{u}_0 + \boldsymbol{\epsilon}$, for every small $\boldsymbol{\epsilon} > 0$. Inserting this into Eq. (A12), expanding the result in a Taylor series and neglecting higher order terms yields for our two-dimensional system

$$\dot{\boldsymbol{\epsilon}} = \mathbf{J}(\boldsymbol{u}_0)\boldsymbol{\epsilon}, \quad \mathbf{J} = \begin{pmatrix} \frac{\partial \dot{n}_s}{\partial n_s} & \frac{\partial \dot{n}_s}{\partial S} \\ \frac{\partial \dot{S}}{\partial n_s} & \frac{\partial \dot{S}}{\partial S} \end{pmatrix}, \tag{A23}$$

with $\mathbf{J}$ the Jacobian matrix. Through eigendecomposition, the characteristic polynomial and with it the eigenvalues of the system can be found. The eigenvalues determine the stability of the fixed points: in the most general case, if $\Re(\sigma_i) < 0$, for all values of i, a fixed point is stable, i.e. attracting, while it is unstable, i.e. repelling, if $\Re(\sigma_i) > 0$, for any i. If $\Re(\sigma_i) = 0$, it is a critical point. For a two-dimensional system, the eigenvalues become

$$\sigma^{\pm} = \frac{1}{2}\mathrm{tr}\mathbf{J} \pm \frac{1}{2}\sqrt{(\mathrm{tr}\mathbf{J})^2 - 4\det\mathbf{J}}, \tag{A24}$$

where tr and det are the trace and determinant of the matrix, respectively. When investigating the stability at fixed point $\boldsymbol{u}_0 = (0, 0.5)$, one obtains

$$\mathbf{J}(\boldsymbol{u}_0) = \begin{pmatrix} -1 & \ell \\ 0 & 1 - E_u + \frac{E_s - E_u}{2} \end{pmatrix}. \tag{A25}$$

The trace and determinant then yield

$$\mathrm{tr} = \frac{E_s - E_u}{2} - E_u, \quad \det = 0 \tag{A26}$$

and

$$\sigma_{S,n_s} = \{0, \frac{E_s - E_u}{2} - E_u\}. \tag{A27}$$

This means that the fixed point is critical along the S-direction ($\sigma_S = 0$) and attractive along the n_s-direction ($\sigma_{n_s} = \frac{E_s - E_u}{2} - E_u < 0$ for $E_u = 1.5$ and $E_s = 0.5$). Trajectories will approach the fixed point along the n_s-direction and its stable manifold (trajectories associated with the fixed point). Along the S-direction the fixed point is approached via a centre manifold: while trajectories are still attracted towards the point, in its proximity they experience a critical slowdown, that is they approach the fixed point increasingly slowly. Along the n_s-direction, the system thus exhibits fast dynamics, while the dynamics along S are slow. This agrees with the assumption that the social dynamics are fast compared to the resource dynamics.

The stability and the location of the fixed point are confirmed when graphically analysing trajectories of the system state in the phase diagram of S and n_s: Fig. A1 shows trajectories with different origins in the phase space and the nullclines. It can be seen that all trajectories trend to the fixed point, making the system outcome always unsustainable since $S = 0$ there.

For different choices of E_u and E_s, the system goes through a transcritical bifurcation along the S-direction. This bifurcation is typical for logistic curves of the shape $\dot{u} = u(\mu - u)$, where μ is the critical parameter. Keeping n_s fixed in Eq. (A10) and analysing the system along the S-direction, one can find $\mu = 1 - E_u + \frac{E_u - E_s}{2}$. For $E_u = 1.5$ and $E_s = 0.5$, $\mu = 0$, resulting in a critical point along the S-direction.

For higher effort levels, the fixed point at $(0, 0.5)$ becomes a stable fixed point, that is the system always becomes unsustainable with an end stock of $S = 0$. For lower effort levels, e.g. $E_u = 1.25$ and $E_s = 0.25$, the critical fixed point turns into a hyperbolic saddle point at $n_{s,0} \cap S_0^1$, having one stable manifold along which trajectories approach it and one unstable manifold along which trajectories are repelled from the point. The other intersect, $n_{s,0} \cap S_0^2$, turns into a stable fixed point moving away from the other fixed point in the phase space.

A2   Comparison to computational model

Figure A2 shows the trajectory of the average model quantities in phase space compared to analytically found trajectories. The model trajectory approaches the fixed point towards the end of the run, following the stable manifold and overlapping with the analytical trajectories. This confirms that the analytical and the actual model have a sufficiently similar outcome and justifies the simplifications made when deducing the former. The sequential colour code indicates that the trajectory quite quickly approaches the fixed point along the $\langle S \rangle$-direction, while the descent towards the analytical fixed point is subject to a critical slowdown. In fact, the drop from

$\langle S \rangle = 1$ to $\langle S \rangle < 0.1$ happens on average in 0.1 %–0.2 % of the run time and the rest of the run time is taken up by the critical slowdown. The critical slowdown in the computational model can be explained by the following effect: $\langle n_s \rangle$ quickly becomes ≈ 0.5, as the difference in the probabilities $p_{s \to u}$ and $p_{u \to s}$ quickly decreases with decreasing stock and individuals become likely to explore both states at the same rate. On average, an individual is then sustainable as long as it is unsustainable, switching between both behaviours. But the timescale to rebuild stock is higher than the timescale to deplete it. This leaves the stock to slowly decrease on average. Additionally, the probability to become unsustainable is increased in comparison to the probability to be sustainable with rising stock, inhibiting agents to leave their unsustainable trajectory. The results of the analytical and the computational investigation suggest that this system, in which agents are socially inert because of economic considerations, ends up in an unsustainable state.
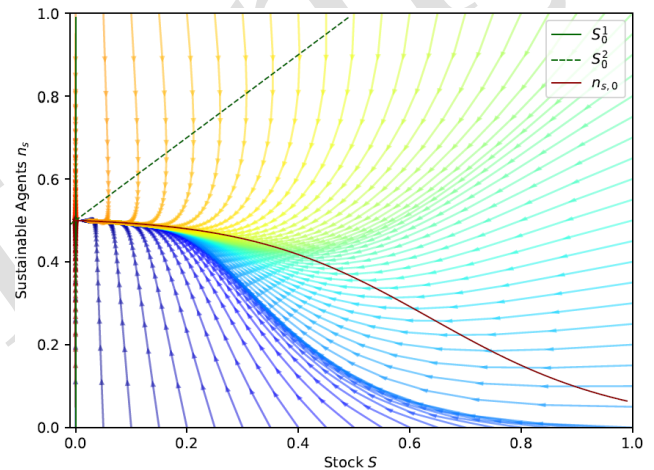
Figure A1. Trajectories of the system state in the phase diagram of the stock S and the fraction of sustainable agents n_s, colour coded according to their origin. The parameter choice is $E_u = 1.5$, $E_s = 0.5$ and $k = 3$. The red curve shows the nullcline of n_s and the green straights show the nullclines of S. All trajectories trend to their intersection, the fixed point u_0.
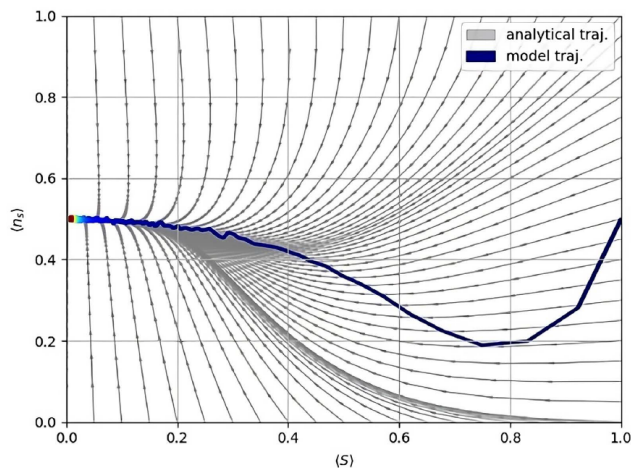
Figure A2. Trajectory of the ensemble average quantities of n_r model runs in the phase diagram of the stock $\langle S \rangle$ and the average fraction of sustainable individuals $\langle n_s \rangle$. The trajectory is colour coded according to the run time, from blue ($t = 0$) to red ($t = t_{\mathrm{end}}$). In the background, trajectories of the analytical model are plotted in grey. The parameter choice is $\Delta T_a = 0.1$, $\Phi_a = 1$, $E_u = 1.5$, $E_s = 0.5$, and $k = 3$.

Code availability. The code of our model implemented in copan:CORE is published on Zenodo https://doi.org/10.5281/zenodo.13220767 (Bechthold et al., 2024) and available at GitHub https://github.com/pik-copan/pycopancore/tree/maxploit (last access: 21 July 2025).

Data availability. No data sets were used in this article. TS5

Author contributions. MB, JH, and JFD conceptualised the social norm framework and the study. MB designed the study, implemented extensions in copan:CORE, performed the simulations, designed the figures of this manuscript, and led the writing of the manuscript with input from all authors. SMC and LS ensured coherence in the social modelling. JB and JH supported the technical implementation of the study. JFD led the supervision of this study. All authors have reviewed and edited the final version of the manuscript. Besides first and last position, the authors are named in alphabetical order.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Earth System Dynamics*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. This work is based on a master's thesis at the University of Heidelberg, conducted externally at PIK's Future-Lab on Earth Resilience in the Anthropocene. The research has been conceived in the scope of the Earth Resilience and Sustainability Initiative (https://www.earthresiliencesustainability.org, last access: 21 July 2025). AI tools have been used for a grammar check on a previous version of this paper. The authors further acknowledge the support of Hannah Prawitz, Niklas Kitzmann, Leander John, Ronja Hotz, and Ricarda Winkelmann during the study.

Financial support. Max Bechthold and Jonathan F. Donges received support from the European Research Council Advanced Grant project ERA (Earth Resilience in the Anthropocene, ERC-2016-ADG-743080). Jonathan F. Donges received financial support from the Federal Ministry for Education and Research (BMBF) (project "PIK Change", grant no. TS6 01LS2001A). Jannes Breier, Luana Schwarz, and Jonathan F. Donges received support from the Generation Foundation TS7, the Global Challenges Foundation, and Partners for a New Economy via the Earth4all project, as well as the European Union's Horizon 2.5 TS8 – Climate Energy and Mobility programme under grant agreement no. 101081661 (project World-Trans). The authors received CE2 support and resources for this project from the European Regional Development Fund (ERDF), the BMBF, and the Land Brandenburg TS9 high-performance computer system at the Potsdam Institute for Climate Impact Research (PIK). TS10

The article processing charges for this open-access publication were covered by the Potsdam Institute for Climate Impact Research (PIK).

Review statement. This paper was edited by James Dyke and reviewed by two anonymous referees.

References

Ajzen, I.: The theory of planned behavior, Organ. Behav. Hum. Dec., 50, 179–211, https://doi.org/10.1016/0749-5978(91)90020-T, 1991.

Ali, Q., Bauch, C. T., and Anand, M.: Coupled Human-Environment Dynamics of Forest Pest Spread and Control in a Multi-Patch, Stochastic Setting, PLOS ONE, 10, e0139353, https://doi.org/10.1371/journal.pone.0139353, 2015.

Anderies, J. M., Barfuss, W., Donges, J. F., Fetzer, I., Heitzig, J., and Rockström, J.: A modeling framework for World-Earth system resilience: Exploring social inequality and Earth system tipping points, Environ. Res. Lett., 18, 095001, https://doi.org/10.1088/1748-9326/ace91d, 2023.

Andreoni, J., Nikiforakis, N., and Siegenthaler, S.: Predicting social tipping and norm change in controlled ex-

periments, P. Natl. Acad. Sci. USA, 118, e2014893118, https://doi.org/10.1073/pnas.2014893118, 2021.

Andrighetto, G. and Vriens, E.: A research agenda for the study of social norm change, Philos. T. R. Soc. A, 380, 20200411, https://doi.org/10.1098/rsta.2020.0411, 2022.

APA: APA Dictionary of Psychology, https://dictionary.apa.org/group (last access: 24 April 2023), 2023.

Axelrod, R.: An Evolutionary Approach to Norms, Am. Polit. Sci. Rev., 80, 1095–1111, https://doi.org/10.2307/1960858, 1986.

Bak-Coleman, J. B., Alfano, M., Barfuss, W., Bergstrom, C. T., Centeno, M. A., Couzin, I. D., Donges, J. F., Galesic, M., Gersick, A. S., Jacquet, J., Kao, A. B., Moran, R. E., Romanczuk, P., Rubenstein, D. I., Tombak, K. J., Bavel, J. J. V., and Weber, E. U.: Stewardship of global collective behavior, P. Natl. Acad. Sci. USA, 118, e2025764118, https://doi.org/10.1073/pnas.2025764118, 2021.

Bandura, A.: Social Learning Theory, Prentice Hall, ISBN 978-0-13-816751-6, 1977.

Barabási, A.-L. and Albert, R.: Emergence of Scaling in Random Networks, Science, 286, 509–512, https://doi.org/10.1126/science.286.5439.509, 1999.

Barfuss, W.: Dynamical systems as a level of cognitive analysis of multi-agent learning: Algorithmic foundations of temporal-difference learning dynamics, Neural Comput. Appl., 34, 1653–1671, 2022.

Barfuss, W., Donges, J. F., Wiedermann, M., and Lucht, W.: Sustainable use of renewable resources in a stylized social–ecological network model under heterogeneous resource distribution, Earth Syst. Dynam., 8, 255–264, https://doi.org/10.5194/esd-8-255-2017, 2017.

Barfuss, W., Donges, J. F., Vasconcelos, V. V., Kurths, J., and Levin, S. A.: Caring for the future can turn tragedy into comedy for long-term collective action under risk of collapse, P. Natl. Acad. Sci. USA, 117, 12915–12922, https://doi.org/10.1073/pnas.1916545117, 2020.

Barlow, L.-A., Cecile, J., Bauch, C. T., and Anand, M.: Modelling Interactions between Forest Pest Invasions and Human Decisions Regarding Firewood Transport Restrictions, PLOS ONE, 9, e90511, https://doi.org/10.1371/journal.pone.0090511, 2014.

Bauch, C. T., Sigdel, R., Pharaon, J., and Anand, M.: Early warning signals of regime shifts in coupled human–environment systems, P. Natl. Acad. Sci. USA, 113, 14560–14567, https://doi.org/10.1073/pnas.1604978113, 2016.

Bechthold, M., Heitzig, J., Donges, J., Barfuss, W., Marc, W., Zimmerer, K. B., Kolb, J., Kolster, T., Müller-Hansen, F., and Breitbach, P.: pycopancore: Reference implementation of the Nexploit model in the copan:CORE World-Earth modelling framework, Zenodo [code], https://doi.org/10.5281/zenodo.13220767, 2024.

Beckage, B., Gross, L., Lacasse, K., Carr, E., Metcalf, S., Winter, J., Howe, P., Fefferman, N., Franck, T., Zia, A., Kinzig, A., and Hoffman, F.: Linking models of human behaviour and climate alters projected climate change, Nat. Clim. Change, 8, 79–84, https://doi.org/10.1038/s41558-017-0031-7, 2018.

Beckage, B., Lacasse, K., Winter, J. M., Gross, L. J., Fefferman, N., Hoffman, F. M., Metcalf, S. S., Franck, T., Carr, E., Zia, A., et al.: The Earth has humans, so why don't our climate models?, Climatic Change, 163, 181–188, 2020.

Bernhard, H., Fehr, E., and Fischbacher, U.: Group Affiliation and Altruistic Norm Enforcement, Am. Econ. Rev., 96, 217–221, https://doi.org/10.1257/000282806777212594, 2006.

Bianchi, F. and Squazzoni, F.: Agent-based models in sociology, WIREs Computational Statistics, 7, 284–306, https://doi.org/10.1002/wics.1356, 2015.

Bicchieri, C.: Norms in the Wild: How to Diagnose, Measure, and Change Social Norms, Oxford University Press, ISBN 9780190622046, https://doi.org/10.1093/acprof:oso/9780190622046.001.0001, 2017.

Biermann, F., Betsill, M. M., Gupta, J., Kanie, N., Lebel, L., Liverman, D., Schroeder, H., Siebenhüner, B., and Zondervan, R.: Earth system governance: a research framework, Int. Environ. Agreem.-P., 10, 277–298, 2010.

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., and Zanin, M.: The structure and dynamics of multilayer networks, Phys. Rep., 544, 1–122, https://doi.org/10.1016/j.physrep.2014.07.001, 2014.

Bonan, J., Cattaneo, C., d'Adda, G., and Tavoni, M.: The interaction of descriptive and injunctive social norms in promoting energy conservation, Nature Energy, 5, 900–909, https://doi.org/10.1038/s41560-020-00719-z, 2020a.

Bonan, J., Cattaneo, C., d'Adda, G., and Tavoni, M.: The interaction of descriptive and injunctive social norms in promoting energy conservation, Nature Energy, 5, 900–909, https://doi.org/10.1038/s41560-020-00719-z, 2020b.

Breier, J., Schwarz, L., Donges, J. F., Gerten, D., and Rockström, J.: Regenerative agriculture for food security and ecological resilience: illustrating global biophysical and social spreading potentials, Potsdam Institute for Climate Impact Research, 16 pp., https://doi.org/10.48485/pik.2023.001, 2023.

Brulle, R. J. and Norgaard, K. M.: Avoiding cultural trauma: climate change and social inertia, Environ. Polit., 28, 886–908, https://doi.org/10.1080/09644016.2018.1562138, 2019.

Bury, T. M., Bauch, C. T., and Anand, M.: Charting pathways to climate change mitigation in a coupled socio-climate model, PLOS Comput. Biol., 15, e1007000, https://doi.org/10.1371/journal.pcbi.1007000, 2019.

Calvin, K. and Bond-Lamberty, B.: Integrated human-earth system modeling – State of the science and future directions, Environ. Res. Lett., 13, 063006, https://doi.org/10.1088/1748-9326/aac642, 2018.

Carpenter, S. R., Folke, C., Scheffer, M., and Westley, F. R.: Dancing on the volcano: social exploration in times of discontent, Ecol. Soc., 24, 23, https://doi.org/10.5751/ES-10839-240123, 2019.

Carrington, M. J., Neville, B. A., and Whitwell, G. J.: Why Ethical Consumers Don't Walk Their Talk: Towards a Framework for Understanding the Gap Between the Ethical Purchase Intentions and Actual Buying Behaviour of Ethically Minded Consumers, J. Bus. Ethics, 97, 139–158, https://doi.org/10.1007/s10551-010-0501-6, 2010.

Castilla-Rho, J. C., Rojas, R., Andersen, M., Holley, C., and Mariethoz, G.: Social tipping points in global groundwater management, Nature Human Behaviour, 1, 640–649, https://doi.org/10.1038/s41562-017-0181-7, 2017.

Centola, D.: The Spread of Behavior in an Online Social Network Experiment, Science, 329, 1194–1197, https://doi.org/10.1126/science.1185231, 2010.

Centola, D.: The Social Origins of Networks and Diffusion, Am. J. Sociol., 120, 1295–1338, https://doi.org/10.1086/681275, 2015.

Centola, D. and Macy, M.: Complex Contagions and the Weakness of Long Ties, Am. J. Sociol., 113, 702–734, https://doi.org/10.1086/521848, 2007.

Centola, D., Willer, R., and Macy, M.: The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms, Am. J. Sociol., 110, 1009–1040, https://doi.org/10.1086/427321, 2005.

Ceschi, A., Sartori, R., Dickert, S., Scalco, A., Tur, E., Tommasi, F., and Delfini, K.: Testing a norm-based policy for waste management: An agent-based modeling simulation on nudging recycling behavior, J. Environ. Manage., 294, 112938, https://doi.org/10.1016/j.jenvman.2021.112938, 2021.

Cialdini, R., Reno, R., and Kallgren, C.: A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places, J. Pers. Soc. Psychol., 58, 1015–1026, https://doi.org/10.1037/0022-3514.58.6.1015, 1990.

Constantino, S. M., Schlüter, M., Weber, E., and Wijermans, N.: Cognition and behavior in context: a framework and theories to explain natural resource use decisions in social-ecological systems, Sustain. Sci., 16, 1651–1671, https://doi.org/10.1007/s11625-021-00989-w, 2021a.

Constantino, S. M., Pianta, S., Rinscheid, A., Frey, R., and Weber, E. U.: The source is the message: the impact of institutional signals on climate change–related norm perceptions and behaviors, Climatic Change, 166, 35, https://doi.org/10.1007/s10584-021-03095-z, 2021b.

Constantino, S. M., Sparkman, G., Kraft-Todd, G., Bicchieri, C., Centola, D., Shell-Duncan, B., Vogt, S., and Weber, E.: Scaling Up Change: A Critical Review and Practical Guide to Harnessing Social Norms for Climate Action, Psychol. Sci. Publ. Int., 23, 50–97, https://doi.org/10.1177/15291006221105279, 2022.

Crameri, F., Shephard, G. E., and Heron, P. J.: The misuse of colour in science communication, Nat. Commun., 11, 5444, https://doi.org/10.1038/s41467-020-19467-7, 2020.

Crutzen, P.: Geology of Mankind, Nature, 415, 23, https://doi.org/10.1038/415023a, 2002.

Davis, C. A., Heiman, J. R., and Menczer, F.: A Role for Network Science in Social Norms Intervention, Procedia Comput. Sci., 51, 2217–2226, https://doi.org/10.1016/j.procs.2015.05.499, 2015.

Donges, J. F., Lucht, W., Müller-Hansen, F., and Steffen, W.: The technosphere in Earth System analysis: A coevolutionary perspective, The Anthropocene Review, 4, 23–33, https://doi.org/10.1177/2053019616676608, 2017a.

Donges, J. F., Winkelmann, R., Lucht, W., Cornell, S. E., Dyke, J. G., Rockström, J., Heitzig, J., and Schellnhuber, H.: Closing the loop: Reconnecting human dynamics to Earth System science, The Anthropocene Review, 4, 151–157, https://doi.org/10.1177/2053019617725537, 2017b.

Donges, J. F., Heitzig, J., Barfuss, W., Wiedermann, M., Kassel, J. A., Kittel, T., Kolb, J. J., Kolster, T., Müller-Hansen, F., Otto, I. M., Zimmerer, K. B., and Lucht, W.: Earth system modeling with endogenous and dynamic human societies: the copan:CORE open World–Earth modeling framework, Earth Syst. Dynam., 11, 395–413, https://doi.org/10.5194/esd-11-395-2020, 2020.

Donges, J. F., Lucht, W., Cornell, S. E., Heitzig, J., Barfuss, W., Lade, S. J., and Schlüter, M.: Taxonomies for structuring models for World–Earth systems analysis of the Anthropocene: subsystems, their interactions and social–ecological feedback loops, Earth Syst. Dynam., 12, 1115–1137, https://doi.org/10.5194/esd-12-1115-2021, 2021.

Ehrlich, P. R. and Levin, S. A.: The Evolution of Norms, PLOS Biol., 3, e194, https://doi.org/10.1371/journal.pbio.0030194, 2005.

Elsenbroich, C. and Gilbert, N.: Modelling Norms, Springer Dordrecht, ISBN 978-94-017-8514-3, https://doi.org/10.1007/978-94-007-7052-2, 2015.

Epstein, J.: Learning to Be Thoughtless: Social Norms and Individual Computation, Comput. Econ., 18, 9–24, https://doi.org/10.1023/A:1013810410243, 2001.

Erdős, P. and Rényi, A.: On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci, 5, 17–60, 1960.

Falk, A., Andre, P., Boneva, T., and Chopra, F.: Fighting Climate Change: The Role of Norms, Preferences, and Moral Values, SSRN Journal, 59 pp., https://doi.org/10.2139/ssrn.3885418, 2021.

Farahbakhsh, I., Bauch, C. T., and Anand, M.: Best response dynamics improve sustainability and equity outcomes in common-pool resources problems, compared to imitation dynamics, J. Theor. Biol., 509, 110476, https://doi.org/10.1016/j.jtbi.2020.110476, 2021.

Farahbakhsh, I., Bauch, C. T., and Anand, M.: Modelling coupled human–environment complexity for the future of the biosphere: strengths, gaps and promising directions, Philos. T. R. Soc. B, 377, 20210382, https://doi.org/10.1098/rstb.2021.0382, 2022.

Finnemore, M. and Sikkink, K.: International Norm Dynamics and Political Change, Int. Organ., 52, 887–917, https://doi.org/10.1162/002081898550789, 1998.

Forsyth, D. R.: Group Dynamics, 6th edn., Cengage Learning, 2018. TS11

Gavrilets, S.: The dynamics of injunctive social norms, Evolutionary Human Sciences, 2, e60, https://doi.org/10.1017/ehs.2020.58, 2020.

Gavrilets, S. and Richerson, P. J.: Collective action and the evolution of social norm internalization, P. Natl. Acad. Sci. USA, 114, 6068–6073, https://doi.org/10.1073/pnas.1703857114, 2017.

Geier, F., Barfuss, W., Wiedermann, M., Kurths, J., and Donges, J. F.: The physics of governance networks: critical transitions in contagion dynamics on multilayer adaptive networks with application to the sustainable use of renewable resources, Eur. Phys. J.-Spec. Top., 228, 2357–2369, https://doi.org/10.1140/epjst/e2019-900120-4, 2019.

Gelfand, M. J., Gavrilets, S., and Nunn, N.: Norm Dynamics: Interdisciplinary Perspectives on Social Norm Emergence, Persistence, and Change, Annu. Rev. Psychol., 75, 341–378, https://doi.org/10.1146/annurev-psych-033020-013319, 2024.

Geschke, D. and Frindte, W.: Henri Tajfel (Hg.): Social Identity and Intergroup Relations, Cambridge University Press: Cambridge 1982, 546 S., in: Klassiker der Sozialwissenschaften, chap. 75, Springer VS Wiesbaden, 325–329, https://doi.org/10.1007/978-3-658-13213-2, 2016.

Gössling, S., Humpe, A., and Bausch, T.: Does "flight shame" affect social norms? Changing perspectives on the desirabil-

ity of air travel in Germany, J. Clean. Prod., 266, 122015, https://doi.org/10.1016/j.jclepro.2020.122015, 2020.

Granovetter, M.: Threshold Models of Collective Behavior, Am. J. Sociol., 83, 1420–1443, https://www.journals.uchicago.edu/doi/abs/10.1086/226707 (last access: TS12), 1978.

Gross, T. and Blasius, B.: Adaptive coevolutionary networks: a review, J. R. Soc. Interface, 5, 259–271, https://doi.org/10.1098/rsif.2007.1229, 2008.

Guilbeault, D., Becker, J., and Centola, D.: Complex Contagions: A Decade in Review, in: Complex Spreading Phenomena in Social Systems, edited by: Lehmann, S. and Ahn, Y.-Y., 3–25, ISBN 978-3-319-77331-5, https://doi.org/10.1007/978-3-319-77332-2_1, 2018.

Heitzig, J., Kittel, T., Donges, J. F., and Molkenthin, N.: Topology of sustainable management of dynamical systems with desirable states: from defining planetary boundaries to safe operating spaces in the Earth system, Earth Syst. Dynam., 7, 21–50, https://doi.org/10.5194/esd-7-21-2016, 2016.

Henrich, J. and Ensminger, J.: Theoretical foundations: The co-evolution of social norms, intrinsic motivation, markets, and the institutions of complex societies, in: Experimenting with social norms: Fairness and punishment in cross-cultural perspective, Russell Sage Foundation, 19–44, ISBN 978-1-61044-840-6, 2014.

Homans, G. C.: The Human Group, Routledge, https://doi.org/10.4324/9781315132518, 2017.

Hunter, S.: If Ever the Twain Shall Meet: Graph Theoretical Dimensions of Formal and Informal Organization Structure, Int. J. Soc. Sci. Stud., 4, 79, TS13, 2016.

Janssen, M.: Modelling social norms of water conservation, Nature Human Behaviour, 1, 624–625, https://doi.org/10.1038/s41562-017-0196-0, 2017.

Kinzig, A., Ehrlich, P., Alston, L., Arrow, K., Barrett, S., Buchman, T., Daily, G., Levin, B., Levin, S. A., and Oppenheimer, M.: Social Norms and Global Environmental Challenges: The Complex Interaction of Behaviors, Values, and Policy, BioScience, 63, 164–175, https://doi.org/10.1525/bio.2013.63.3.5, 2013.

Lade, S., Tavoni, A., Levin, S. A., and Schlüter, M.: Regime shifts in a social-ecological system, Theor. Ecol., https://doi.org/10.1007/s12080-013-0187-3, 2013.

Legros, S. and Cislaghi, B.: Mapping the Social-Norms Literature: An Overview of Reviews, Perspect. Psychol. Sci., 15, 62–80, https://doi.org/10.1177/1745691619866455, 2019.

Lenton, T. M. and Latour, B.: Gaia 2.0, Science, 361, 1066–1068, https://doi.org/10.1126/science.aau0427, 2018.

Li, N., Hilgard, J., Scheufele, D. A., Winneg, K. M., and Jamieson, K. H.: Cross-pressuring conservative Catholics? Effects of Pope Francis' encyclical on the U.S. public opinion on climate change, Climatic Change, 139, 367–380, https://doi.org/10.1007/s10584-016-1821-z, 2016.

Lin, C.-Y., Yang, Y. C. E., Malek, K., and Adam, J. C.: An investigation of coupled natural human systems using a two-way coupled agent-based modeling framework, Environ. Modell. Softw., 155, 105451, https://doi.org/10.1016/j.envsoft.2022.105451, 2022.

Liu, H., Alharthi, M., Atil, A., Zafar, M., and Khan, I.: A non-linear analysis of the impacts of natural resources and education on environmental quality: Green energy and its role in the future, Resour. Policy, 79, 102940, https://doi.org/10.1016/j.resourpol.2022.102940, 2022.

Malhi, Y., Roberts, J. T., Betts, R. A., Killeen, T. J., Li, W., and Nobre, C. A.: Climate Change, Deforestation, and the Fate of the Amazon, Science, 319, 169–172, https://doi.org/10.1126/science.1146961, 2008.

Maurstad, A.: To Fish or Not to Fish: Small-Scale Fishing and Changing Regulations of the Cod Fishery in Northern Norway, Hum. Organ., 59, 37–47, https://doi.org/10.17730/humo.59.1.q0242m112x223862, 2000.

McDonald, R. I. and Crandall, C. S.: Social norms and social influence, Current Opinion in Behavioral Sciences, 3, 147–151, https://doi.org/10.1016/j.cobeha.2015.04.006, 2015.

McPherson, M., Smith-Lovin, L., and Cook, J. M.: Birds of a Feather: Homophily in Social Networks, Annu. Rev. Sociol., 27, 415–444, https://doi.org/10.1146/annurev.soc.27.1.415, 2001.

Müller-Hansen, F., Schlüter, M., Mäs, M., Donges, J. F., Kolb, J. J., Thonicke, K., and Heitzig, J.: Towards representing human behavior and decision making in Earth system models – an overview of techniques and approaches, Earth Syst. Dynam., 8, 977–1007, https://doi.org/10.5194/esd-8-977-2017, 2017.

Neidhardt, F.: Das innere System sozialer Gruppen, Kölner Z. Soziol. Soz., 69, 433–454, https://doi.org/10.1007/s11577-017-0415-8, 2017.

Newman, M.: Social networks, in: Networks, Oxford University Press, ISBN 9780198805090, https://doi.org/10.1093/oso/9780198805090.003.0004, 2018.

Nhim, T., Richter, A., and Zhu, X.: The resilience of social norms of cooperation under resource scarcity and inequality – An agent-based model on sharing water over two harvesting seasons, Ecol. Complex., 40, 100709, https://doi.org/10.1016/j.ecocom.2018.06.001, publisher: Elsevier, 2019.

Nitzbon, J., Heitzig, J., and Parlitz, U.: Sustainability, collapse and oscillations in a simple World-Earth model, Environ. Res. Lett., 12, 074020, https://doi.org/10.1088/1748-9326/aa7581, 2017.

Nowak, M. and Sigmund, K.: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game, Nature, 364, 56–58, https://doi.org/10.1038/364056a0, 1993.

Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W. N., Arrow, K. J., Barrett, S., Carpenter, S., Chapin, F. S., Crépin, A., Daily, G., Ehrlich, P., Folke, C., Jager, W., Kautsky, N., Levin, S. A., Madsen, O. J., Polasky, S., Scheffer, M., Walker, B., Weber, E. U., Wilen, J., Xepapadeas, A., and de Zeeuw, A.: Social norms as solutions, Science, 354, 42–43, https://doi.org/10.1126/science.aaf8317, 2016.

Nøstbakken, L.: Formal and informal quota enforcement, Resour. Energy Econ., 35, 191–215, https://doi.org/10.1016/j.reseneeco.2012.10.001, 2013.

Olson, M.: The Logic of Collective Action: Public Goods and the Theory of Groups, Second Printing with a New Preface and Appendix, Harvard University Press, ISBN 9780674537507, https://doi.org/10.2307/j.ctvjsf3ts, 1971.

Opp, K.: How do norms emerge? An outline of a theory, Mind & Society, 2, 101–128, https://doi.org/10.1007/BF02512077, 2001.

Ostrom, E.: Collective Action and the Evolution of Social Norms, J. Econ. Perspect., 14, 137–158, https://doi.org/10.1257/jep.14.3.137, 2000.

Otto, I. M., Donges, J. F., Cremades, R., Bhowmik, A., Hewitt, R. J., Lucht, W., Rockström, J., Allerberger, F., McCaffrey, M., Doe, S.

S. P., Lenferna, A., Morán, N., van Vuuren, D. P., and Schellnhuber, H.: Social tipping dynamics for stabilizing Earth's climate by 2050, Proceedings of the National Acad. Sci. USA, 117, 2354–2365, https://doi.org/10.1073/pnas.1900577117, 2020.

Perman, R.: Natural Resource and Environmental Economics, Pearson Education, ISBN 978-0-273-65559-6, 2003. TS14

Reed, M. S., Evely, A. C., Cundill, G., Fazey, I., Glass, J., Laing, A., Newig, J., Parrish, B., Prell, C., Raymond, C., and Stringer, L. C.: What is Social Learning?, Ecol. Soc., 15, https://www.jstor.org/stable/26268235 (last access: 21 July 2025), 2010.

Restrepo-Plaza, L. and Fatas, E.: When ingroup favoritism is not the social norm a lab-in-the-field experiment with victims and non-victims of conflict in Colombia, J. Econ. Behav. Organ., 194, 363–383, https://doi.org/10.1016/j.jebo.2021.12.025, 2022.

Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin III, F. S., Lambin, E., Lenton, T., Scheffer, M., Folke, C., and Schellnhuber, H.: A Safe Operating Space for Humanity, Nature, 461, 472–475, 2009.

Roth-Fauchere, K.: Chaos – Complexity – Evolution, https://arolla.net/werkstatt/science/chaos-complexity-evolution (last access: 24 May 2023), 2023.

Saaty, T. L.: Résumé of Useful Formulas in Queuing Theory, Oper. Res., 5, 161–200, 1957.

Satake, A., Leslie, H., Iwasa, Y., and Levin, S. A.: Coupled ecological–social dynamics in a forested landscape: Spatial interactions and information flow, J. Theor. Biol., 246, 695–707, https://doi.org/10.1016/j.jtbi.2007.01.014, 2007.

Schelling, T.: Micromotives and Macrobehavior, Norton, ISBN 0393090094, 1978.

Schelling, T. C.: Dynamic models of segregation, J. Math. Sociol., 1, 143–186, https://doi.org/10.1080/0022250X.1971.9989794, 1971.

Schellnhuber, H.: "Earth system" analysis and the second Copernican revolution, Nature, 402, C19–C23, 1999.

Schellnhuber, H.: Tipping elements in the Earth System, P. Natl. Acad. Sci. USA, 106, 20561–20563, https://doi.org/10.1073/pnas.0911106106, 2009.

Schlüter, M., Mcallister, R. R. J., Arlinghaus, R., Bunnefeld, N., Eisenack, K., Hölker, F., Milner-Gulland, E., Müller, B., Nicholson, E., Quaas, M., and Stoeven, M.: New Horizons for Managing the Environment: A Review of Coupled Social-Ecological Systems Modeling. Natural Resource Modeling, Nat. Resour. Model., 25, 219–272, https://doi.org/10.1111/j.1939-7445.2011.00108.x, 2012.

Schlüter, M., Tavoni, A., and Levin, S. A.: Robustness of norm-driven cooperation in the commons, P. Roy. Soc. B-Biol. Sci., 283, 20152431, https://doi.org/10.1098/rspb.2015.2431, 2016.

Schultz, P., Nolan, J., Cialdini, R., Goldstein, N., and Griskevicius, V.: The Constructive, Destructive, and Reconstructive Power of Social Norms, Psychol. Sci., 18, 429–34, https://doi.org/10.1111/j.1467-9280.2007.01917.x, 2007.

Sherif, M. and Sherif, C. W.: Research on intergroup relations, in: Perspectives in Social Psychology, edited by: Klineberg, O. and Christie, R., Holt, Rinehart and Winston, 153–177, 1965. TS15

Sigdel, R., Anand, M., and Bauch, C. T.: Competition between injunctive social norms and conservation priorities gives rise to complex dynamics in a model of forest growth and opinion dynamics, J. Theor. Biol., 432, 132–140, https://doi.org/10.1016/j.jtbi.2017.07.029, 2017.

Smith, J. R., Louis, W. R., Terry, D. J., Greenaway, K. H., Clarke, M. R., and Cheng, X.: Congruent or conflicted? The impact of injunctive and descriptive norms on environmental intentions, J. Environ. Psychol., 32, 353–361, https://doi.org/10.1016/j.jenvp.2012.06.001, 2012.

Spears, R.: Social Influence and Group Identity, Annu. Rev. Psychol., 72, 367–390, https://doi.org/10.1146/annurev-psych-070620-111818, 2021.

Staunton, M., Louis, W. R., Smith, J. R., Terry, D. J., and McDonald, R. I.: How negative descriptive norms for healthy eating undermine the effects of positive injunctive norms, J. Appl. Soc. Psychol., 44, 319–330, https://doi.org/10.1111/jasp.12223, 2014.

Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., and Ludwig, C.: The Trajectory of the Anthropocene: The Great Acceleration, The Anthropocene Review, https://doi.org/10.1177/2053019614564785, 2015.

Steffen, W., Richardson, K., Rockström, J., Schellnhuber, H., Dube, O. P., Dutreuil, S., Lenton, T., and Lubchenco, J.: The emergence and evolution of Earth System Science, Nature Reviews Earth & Environment, 1, 54–63, https://doi.org/10.1038/s43017-019-0005-6, 2020.

Suzuki, Y. and Iwasa, Y.: Conflict between groups of players in coupled socio-economic and ecological dynamics, Ecol. Econ., 68, 1106–1115, https://doi.org/10.1016/j.ecolecon.2008.07.024, 2009.

Tavoni, A., Schlüter, M., and Levin, S. A.: The survival of the conformist: Social pressure and renewable resource management, J. Theor. Biol., 299, 152–161, https://doi.org/10.1016/j.jtbi.2011.07.003, 2012.

Thampi, V. A., Anand, M., and Bauch, C. T.: Socio-ecological dynamics of Caribbean coral reef ecosystems and conservation opinion propagation, Sci. Rep., 8, 2597, https://doi.org/10.1038/s41598-018-20341-0, 2018.

Tichy, N. and Fombrun, C.: Network Analysis in Organizational Settings, Hum. Relat., 32, 923–965, https://doi.org/10.1177/001872677903201103, 1979.

Tilman, A., Watson, J., and Levin, S. A.: Maintaining cooperation in social-ecological systems: Effective bottom-up management often requires sub-optimal resource use, Theor. Ecol., 10, 155–165, https://doi.org/10.1007/s12080-016-0318-8, 2016.

Tomasello, M.: Becoming Human: A Theory of Ontogeny, Harvard University Press, ISBN 9780674980853, https://doi.org/10.2307/j.ctv24tr9w1, 2019.

Traulsen, A., Semmann, D., Sommerfeld, R., Krambeck, H.-J., and Milinski, M.: Human strategy updating in evolutionary games, P. Natl. Acad. Sci. USA, 107, 2962–2966, https://doi.org/10.1073/pnas.0912515107, 2010.

Tu, C., Wu, Y., Chen, R., Fan, Y., and Yang, Y.: Balancing Resource and Strategy: Coevolution for Sustainable Common-Pool Resource Management, Earth Systems and Environment, 9, 1529–1542, https://doi.org/10.1007/s41748-024-00489-8, 2024.

Turner, J. C.: Towards a cognitive redefinition of the social group, in: Social Identity and Intergroup Relation, edited by: Tajfel, H., chap. 1, Cambridge University Press, 15–36, ISBN 978-0-521-15365-2, 2010.

Verburg, P., Dearing, J., Dyke, J., Van der Leeuw, S., Seitzinger, S., Steffen, W., and Syvitski, J.: Methods and approaches to modelling the Anthropocene, Global Environmental Change,

39, 328–340, https://doi.org/10.1016/j.gloenvcha.2015.08.007, 2015.

Voss, T.: Game-theoretical perspectives on the emergence of social norms, in: Social Norms, 105–136, ISBN 978-1-61044-280-0, 2005.

Wiedermann, M., Donges, J. F., Heitzig, J., Lucht, W., and Kurths, J.: Macroscopic description of complex adaptive networks co-evolving with dynamic node states, Phys. Rev. E, 91, 052801, https://doi.org/10.1103/PhysRevE.91.052801, 2015.

Winkelmann, R., Donges, J. F., Smith, E. K., Milkoreit, M., Eder, C., Heitzig, J., Katsanidou, A., Wiedermann, M., Wunderling, N., and Lenton, T. M.: Social tipping processes towards climate action: A conceptual framework, Ecol. Econ., 192, 107242, https://doi.org/10.1016/j.ecolecon.2021.107242, 2022.

Young, H.: The Evolution of Social Norms, Annu. Rev. Econ., 7, 359–387, https://doi.org/10.1146/annurev-economics-080614-115322, 2015.

Remarks from the language copy-editor

CE1   Please check and confirm the placement of "middle", "left", and "right".

CE2   Please note the slight edits.

Remarks from the typesetter

TS1   Please check all affiliations and affiliation codes carefully and confirm if they are correct

TS2   Please confirm $N \cdot p$.

TS3   Please give an explanation of why Eq. (5) needs to be changed. We have to ask the handling editor for approval. Thanks.

TS4   Please confirm all vectors/matrices.

TS5   Please confirm.

TS6   Please confirm "grant no.".

TS7   Please confirm this is "Generation Foundation (USA)".

TS8   "Horizon 2020"?

TS9   Please confirm this is "Ministerium für Ländliche Entwicklung, Umwelt und Landwirtschaft des Landes Brandenburg".

TS10   Please confirm both Acknowledgements and Financial support sections. Please also confirm the updated funding information in the MS records.

TS11   Please provide ISBN/DOI.

TS12   Please provide date of last access.

TS13   Please provide DOI.

TS14   Please confirm reference list entry.

TS15   Please provide URL/DOI.