Manuscript prepared for Earth Syst. Dynam.
with version 2015/09/17 7.94 Copernicus papers of the LaTeX class copernicus.cls.
Date: 9 November 2015

Topology of sustainable management of dynamical systems with desirable states: from defining planetary boundaries to safe operating spaces in the Earth System

Jobst Heitzig[1], **Tim Kittel**[1,2], **Jonathan F. Donges**[1,3], **and Nora Molkenthin**[4]

[1]Research Domains Transdisciplinary Concepts & Methods and Earth System Analysis, Potsdam Institute for Climate Impact Research, PO Box 60 12 13, 14412 Potsdam, Germany, EU
[2]Department of Physics, Humboldt University, Newtonstr. 15, 12489 Berlin, Germany, EU
[3]Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, 114 19 Stockholm, Sweden, EU
[4]Department for Nonlinear Dynamics & and Network Dynamics Group, Max Planck Institute for Dynamics and Self-Organization, Bunsenstraße 10, 37073 Göttingen, Germany, EU

Correspondence to: Jobst Heitzig (heitzig@pik-potsdam.de)

Abstract.

To keep the Earth System in a desirable region of its state space, such as defined by the recently suggested "tolerable environment and development window", "guardrails", "planetary boundaries", or "safe (and just) operating space for humanity", one not only needs to understand the quantitative internal dynamics of the system and the available options for influencing it (management), but also the structure of the system's state space with regard to certain qualitative differences. Important questions are: Which state space regions can be reached from which others with or without leaving the desirable region? Which regions are in a variety of senses "safe" to stay in when management options might break away, and which qualitative decision problems may occur as a consequence of this topological structure?

In this article, we develop a mathematical theory of the qualitative topology of the state space of a dynamical system with management options and desirable states, as a complement to the existing literature on optimal control which is more focussed on quantitative optimization and is much applied in both the engineering and the integrated assessment literature. We suggest a certain terminology for the various resulting regions of the state space and perform a detailed formal classification of the possible states with respect to the possibility of avoiding or leaving the undesired region. Our results indicate that before performing some form of quantitative optimization such as of indicators of human well-being for achieving certain sustainable development goals, a sustainable and resilient management of the Earth System may require decisions of a more discrete type that come in the form of several dilemmas, e.g., choosing between eventual safety and uninterrupted desirability, or between uninterrupted safety and larger flexibility.

We illustrate the concepts and dilemmas drawing on conceptual models from climate science, ecology, coevolutionary Earth System modeling, economics, and classical mechanics, and discuss their potential relevance for the climate and sustainability debate, in particular suggesting several levels of planetary boundaries of qualitatively increasing safety.

1 Introduction

The sustainable management of systems mainly governed by an internal dynamics for which one desires to stay in a certain region of their state space, such as a “tolerable environment & development (E&D) window” or within “guardrails” in a model of the Earth System (Schellnhuber, 1998; Petschel-Held et al., 1999; Bruckner and Zickfeld, 2008), requires first and foremost an understanding of the *topology* of the system’s state space in terms of what regions are in some sense “safe” to stay in, and to what qualitative degree, and which of these regions can be reached with some degree of safety from which other regions, either by the internal (“default”) dynamics or by some alternative dynamics influenced by some form of management. In the context of Earth System analysis for studying anthropogenic climate change (Schellnhuber, 1998, 1999), management options may correspond to global climate policies for mitigation of greenhouse gas emissions (Edenhofer et al., 2014) or technological interventions such as geoengineering (Vaughan and Lenton, 2011) and much debated criteria for desirability include the resemblance of a Holocene-like state or the provision of certain levels of human well-being. In this setting, it may be very hard to advance the definition of meaningful “planetary boundaries” and a corresponding “safe operating space for humanity” (Rockström et al., 2009a; Steffen et al., 2015) and relate them to sustainable development goals without such an in-depth analysis.

Also the question whether it suffices to influence the system by active management for only a limited time to reach a safe region or whether it might be necessary to repeat active management indefinitely or even continue it uninterruptedly in order to avoid undesired state space regions, which is closely related to the “sustainability paradigms” of Schellnhuber (1998), seems quite relevant in view of urgent problems such as the climate policy debate. E.g., if suitable climate-change mitigation policies such as certain forms of energy market regulation can transform the economic system in a way that allows one to eventually deregulate the market again, then for how long can one delay mitigation until this feature is lost and only permanent regulation can help? Or, if certain adaptation or geoengineering options might be cheaper than mitigation but require an uninterrupted management or lead to a less well-known region of state space (Kleidon and Renner, 2013), which of these qualitatively different properties is preferable?

We will see that such questions about a “safe” or “safe and just operating space” (Rockström et al., 2009b; Raworth, 2012; Scheffer et al., 2015; Carpenter et al., 2015) may lead to decision dilemmas that cannot as easily be analysed in a purely optimization-based framework, but that are highly relevant for the design of resilient Earth System management strategies. A summary of these dilemmas is contained in Table 1 (the possible examples from Earth System management mentioned there are discussed in the next section).

The paradigm of optimal control, which is much applied in both the engineering, on the one hand does not provide sufficient concepts for such a qualitative analysis and on the other hand typically requires quite a lot of additional knowledge, in particular, some or other form of *quantitative* evaluation of states, e.g., in terms of indicators of human well-being. Of course, the integrated assessment literature, although also using optimization as a basic tool, has realized since long that the spatiotemporal distribution of wealth and the diversity and uncertainty of impacts imply that the problem is hard to frame in terms of a single objective function and has used several techniques to deal with this multi-issue multi-agent decision problem, including certainty-equivalent discount rates and hyperbolic discounting (Dasgupta, 2008), cost-efficiency instead of cost-benefit analyses (Edenhofer et al., 2010), lexicographic preferences (Ayres et al., 2001), and many-objective decision making (Singh et al., 2015), to name only a few, but although qualitative constraints appear in many of them, the actual analyses then typically still focus on quantitative assessments.

In this article, we will complement the above-mentioned set of assessment tools by deriving in a purely topological way a thorough and precise *qualitative* classification of the possible states of a system with respect to the possibility of avoiding or leaving some given undesired region by means of some given management options. Our results indicate that in addition to (or maybe rather before) performing some form of quantitative (constrained) optimization, the sustainable and resilient management of a system may require decisions of a more discrete type, e.g., choosing between eventual safety and permanent desirability, or between permanent safety and increasing future options, etc. This appears even more so in the presence of strong nonlinearities, multistable regimes, bifurcations, and tipping elements (Lenton et al., 2008; Schellnhuber, 2009; Keller et al., 2005), where small state changes due to random perturbations or deliberate management may not only have large consequences but can lead to qualitative and possibly irreversible changes.

To indicate the wide scope of applicability of our concepts in various subdisciplines of Earth System Science, we illustrate the concepts and dilemmas with conceptual models from climate science, ecology, coevolutionary Earth System modeling, economics, and classical mechanics.

In contrast to the somewhat related but more formal approach of sequential decision problems in discrete-time systems (Botta et al., 2015), we focus on the more easily applicable class of *continuous-time* systems and their models here. Our classification is based on a distinction between default and alternative trajectories of a system, and suitably adapted *reachability* concepts from control theory and the important but vast field of viability theory (Aubin, 2009; Aubin et al., 2011; Aubin and Saint-Pierre, 2007; Martin, 2004; Rougé et al., 2013; Frankowska and Quincampoix, 1990). Since physical models of global-scale processes or other macroscopic systems are usually of a statistical physics

Table 1. Preview of dilemma types discussed in the article.

Name	Option 1	Option 2	Possible example
“Glade” dilemma	higher desirability/flexibility	safety	adaptation/mitigation
“Lake” dilemma	uninterrupted desirability	eventual safety	great transformation
“Port” dilemma	higher flexibility	higher desirability	land-use change
“Harbour” dilemma	uninterrupted desirability	eventually higher desirability/flexibility	space colonization
“Dock” dilemma	uninterrupted safety	eventually higher desirability/flexibility	new technologies

nature in the sense that they represent the aggregate effects of many micro-scale processes by suitable approximations, their proper interpretation typically requires one to expect small (actually or seemingly) random perturbations. We take this into account here by strengthening the usual notion of reachability to one of *stable reachability,* and by requiring the featured subsets of state space to be topologically open (instead of closed) sets, so that infinitesimal perturbations cannot kick the system out of them.

In the next subsection (Sec. 1.1), we will briefly summarize our main concepts with the help of a metaphorical illustration, before introducing the corresponding formal notation in Sec. 2 in a concise way, reserving a more detailed formal treatment for Appendix A. The framework is then exemplified at the hand of several low-dimensional, conceptual models from various subdisciplines of Earth System Science including climate science, ecology, and coevolutionary social-environmental Earth System modelling (Sec. 3) in order to indicate the wide scope of applicability of our concepts. A thorough analysis of more realistic and thus higher-dimensional models of the Earth System we have to leave for future studies since that would require further improvement of the numerical methods and algorithms employed for finding region boundaries. We conclude with a discussion and outlook in Sec. 4.

1.1 Metaphorical framework

As a start, let’s take the common metaphor that “we’re all in the same boat” literally and represent the state of the Earth System with all its natural and socio-economic parts at each point in time by a single small boat floating or being rowed somewhere on a rather complex system of waters such as in Fig. 1.

The boat can only be on water, not on land, will generally float along with the stream that represents the inherent dynamics of the Earth System over hundreds and thousands of years (the “default trajectory”), but may also be rowed in more or less different directions depending on how strong the surge of the stream is, and this possibility of rowing represents mankind’s agency in deliberately influencing the Earth System’s course to some extent by some or other form of what we will call “management” below. Let us assume that the main qualitative distinction with regard to where humanity wants their boat to be is represented by a division of the whole region into a desirable, “sunny” region on the left and an undesirable, “dark” region on the right, both containing several parts of the waters that may be connected in any imaginable ways, and with the natural water flow possibly drawing the boat back and forth between these two regions. The sunny region is meant to consist of all those possible states of the natural and socio-economic parts of the Earth System in which some generally agreed environmental and living standards are met, such as those defined by the human rights charter or the sustainable development goals (global goals) recently adopted by the United Nations. An alternative definition of the sunny region has been put forward in the planetary boundary framework (Rockström et al., 2009a; Steffen et al., 2015), where states lying within the corridor of Earth System variability during the Holocene that human societies are adapted to are considered as desirable.

We will show in this article that in such a setting, no matter how the waters look exactly, the general situation is in a certain sense always equivalent to the situation depicted in Fig. 1. There will in general be a certain sunny water region where one does not need to row at all in order to stay in the sun forever but can simply lean back and let the boat float around inside that region. In the picture, this region is the top-left tranquil tarn, but in general this region may also consist of several disconnected parts which we will call the *shelters* to emphasize their desirable and safe nature. Indeed, we will argue below that these shelters may be the most natural candidates for being called a “safe and just operating space for humanity”, only that we may not yet be in them. In the Earth System, there may be several such shelters, one of which might correspond to resilient states of the world (Folke et al., 2010) where humanity lives reconnected to the biosphere (Folke et al., 2011) and no active intervention or constant large-scale management is needed.

Connected to the shelter(s), there will in general also be other parts of the sunny region where it would not be safe to just lean back since the flow would then draw the boat into the dark after some time, but from where the shelters can still be reached by some suitable rowing, as show to the left of the “danger” sign in the image. For their “almost-safe” character, we will call such regions *glades.* If the glade is for

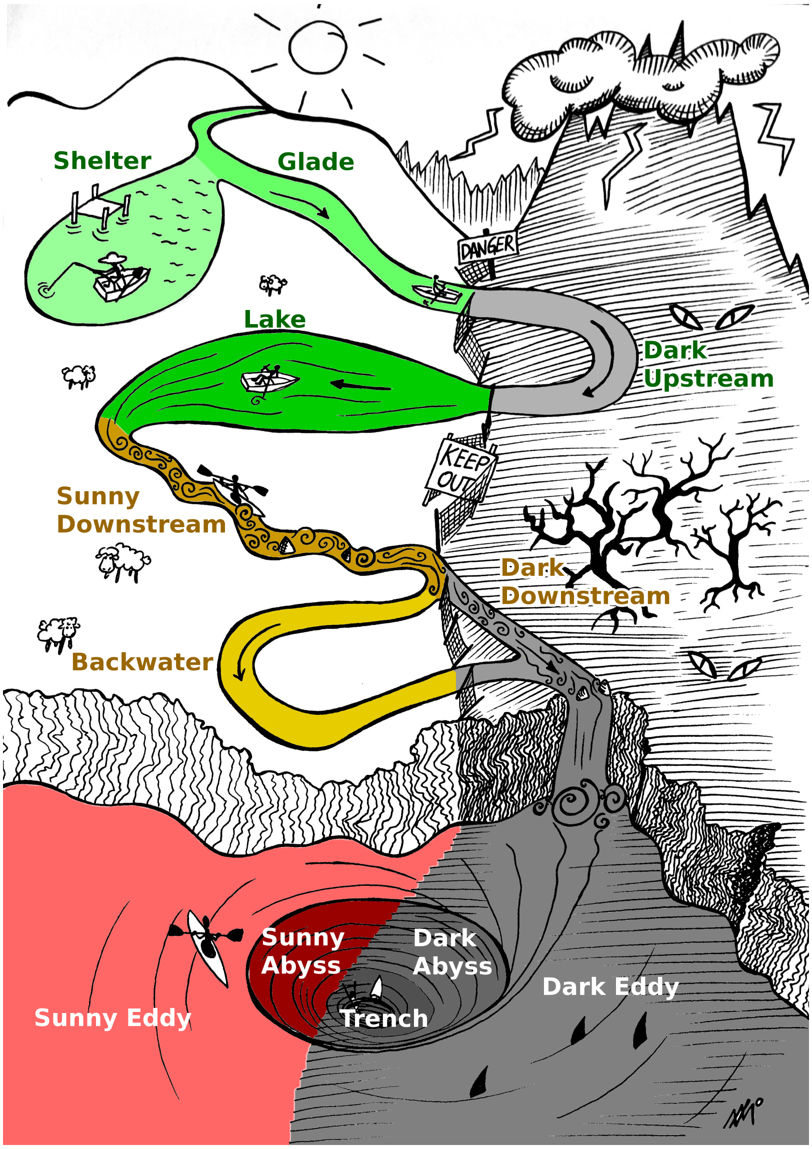

Figure 1. Metaphorical summary of concepts introduced in Section 1.1 inspired by Schellnhuber (1998). It depicts a river flowing from the mountains to the sea while going through sunny (left) and dark parts (right) where humanity can float and row on a boat. In the *shelter*, no rowing is needed to remain in the sun. One can row against the stream direction in slowly flowing parts, shown with long thin arrows, but in fast parts marked with swirls this is not possible. This setting gives rise to a number of qualitatively different regions of the system's state space that can be found in any manageable dynamical system as well: *upstream* regions such as *glades* and *lakes* from where the shelter can be reached, *downstream* regions such as the *backwaters* from where one can at best stay in the sun by management, and several types of worse regions, all labeled here and explained in the text. See also Figs. 2 and 3.

some reason more desirable or offers more flexibility in terms of where one may row, one may face a *dilemma* when in a glade, i.e., a qualitative decision problem, namely whether to prefer staying in the safety of the shelter or in the more desirable but unsafe glade.

The shelters may also be reached by rowing from some places within the dark region (e.g., to the right of the "danger" sign) or through such a dark region from some other sunny places (such as those above the "keep out" sign). Among these latter sunny places from where the shelters can be reached only through the dark, there will generally be some places where one may alternatively stay forever in the sun by continuous rowing instead of passing through the dark and leaning back eventually. Such special places such as the one above the "keep out" sign will be called *lakes* here, and they are characterized by a moderate surge towards a dark place that one can row against and by the decision dilemma

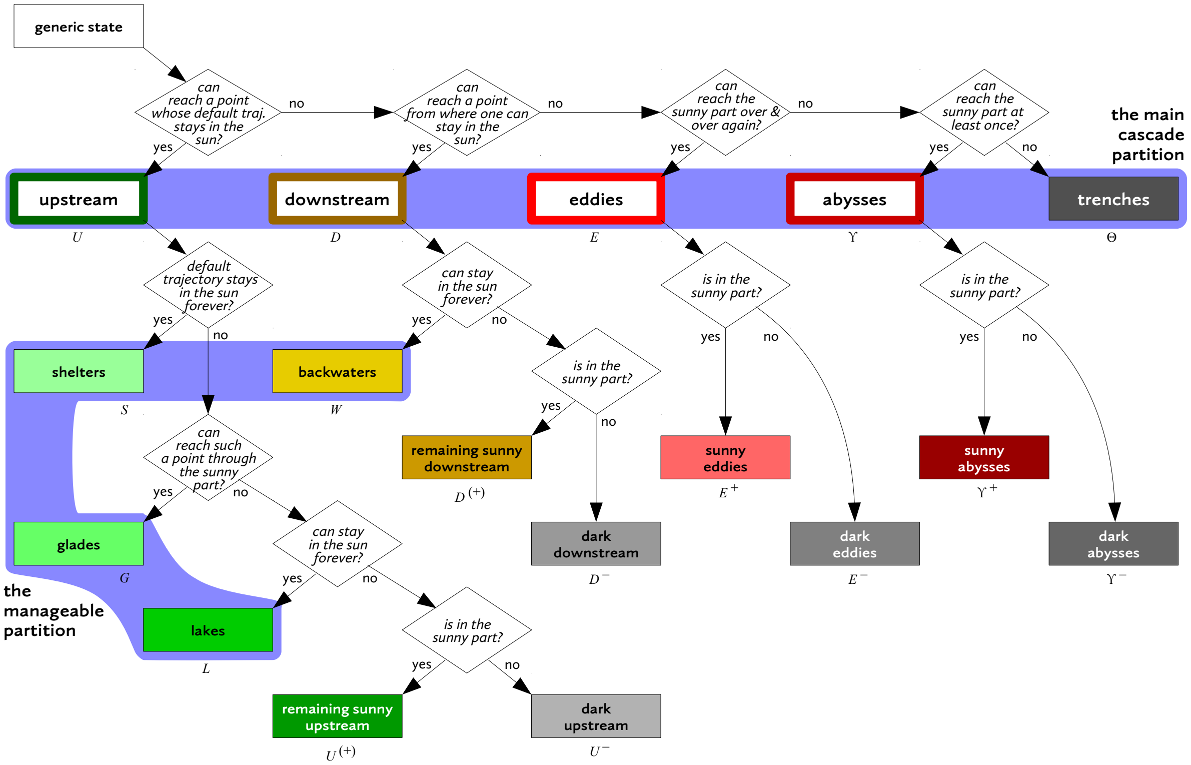

Figure 2. Decision tree summarizing the partition of a manageable dynamical system's state space w.r.t. stable reachability of the desired region or the shelters (main cascade), and the finer partition of the manageable region. The color scheme (grey undesired regions, green upstream regions, yellow downstream regions, red eddies and abysses, lighter meaning better) is also used in the remaining figures.

that results from the question of whether one should indeed do so or rather row to a shelter through the dark.

All these regions together will be called the *upstream* region for reasons that should become clear soon. In any system's state space, the upstream consists of all states from which the shelters can be reached by management, and it is partitioned into one or several shelters, glades, dark upstream parts, lakes, and some remaining sunny upstream parts where it is not possible to stay in the sun forever. In Fig. 1, the upstream ends where the *rapids* left of the "keep out" sign begin since there the stream becomes so strong that it gets impossible to row against it to eventually reach a shelter. Once the boat has left the upstream via such a rapid, there is no hope of leaning back eventually and stay in the sun, and for this reason the borders of the upstream may be called the "no regrets planetary boundaries", forming a middle level of a hierarchy of planetary boundaries we will suggest in Sec. 4.

Further down the stream there will typically be places where it is still possible to stay in the sun forever, only that one has to row over and over again to do so, such as in the slow-moving side branch below the "keep out" sign in the picture. Such regions, called *backwaters* here, are similar to lakes, only without the option of rowing to a shelter, so that the lake dilemma does not occur since the only chance one has is to row against the slow surge to stay in the backwater. While the upstream was defined by being able to reach a shelter, the *downstream* is now defined as all places from where a backwater but not a shelter can be reached, including the backwaters, some dark parts such as the slow moving dark part just right of the backwater in the picture, and maybe some remaining sunny downstream parts from where one may reach a backwater only through the dark. An example for a backwater could be a "machine world" where humanity can fully control nature to its very minute detail. While they can stay within the sunny region for infinite time by this management, there is no way of reaching a shelter anymore because the ecosystem has been changed irreversibly.

The waterfall in Fig. 1 indicates that besides the upstream and downstream regions, where it is possible to stay in the sun eventually, there will in general be further, less hopeful places the system may be in, from where one cannot avoid entering the dark over and over again. In some of those, one can at least make sure that one also spends some time in the sun over and over again, as depicted by the kayak in the picture. Since this is typically connected to some form of cyclic motion, we will call such regions *eddies*. In some eddies, fail-

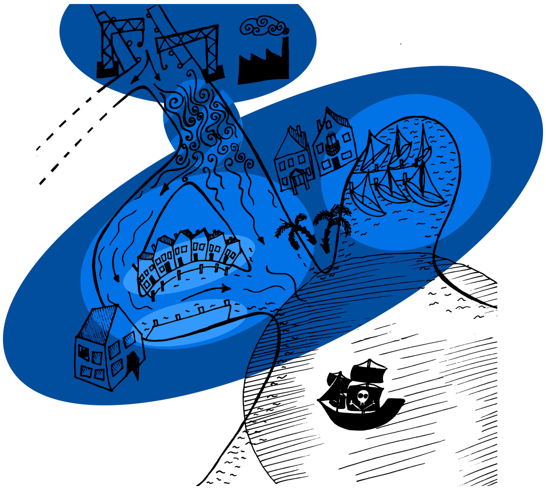

Figure 3. Illustration of port, harbour and dock dilemmas introduced in Section 1.1. As in Fig. 1, humanity can float in and row a boat on a complex waterway. From the upper *port* city (upper dark blue region), one can get to some unknown region to the left and to another, nicer port city (lower dark blue) at the shore through a rapid (hatched blue) which cannot be traversed in the other direction. This choice between desirability and flexibility forms a *port dilemma.* The nicer port city has two harbours (middle blue regions), of which the right one is more desirable, and between which one can switch only through an undesired region where pirates loom (circular area). Boats in the left harbour face the *harbour dilemma* of choosing between either avoiding the undesired region by all means or eventually reaching a place of higher desirability. Finally, in the left harbour there are two safe *docks* (light blue regions), of which the top one is more desirable, and between which one can switch only through an unsafe part of the harbour from which one may be drawn into the undesired region if the engine fails. Boats in the bottom dock face the *dock dilemma* of choosing between uninterrupted safety and eventual higher desirability.

ing to row correctly may push the boat into an even less desirable region, called an *abyss,* from where one can no longer avoid ending up in the dark forever eventually, as in the ring-shaped abyss shown inside the eddy in the figure. Finally, the dark region from where there is no escape, depicted in the center of the abyss, will be called a *trench.*

This completes our main partitioning of the Earth System's or any other manageable system's state space into qualitatively different regions: Upstream and downstream defined by being able to reach shelters or backwaters, abysses defined by not being able to avoid ending up in a trench, and eddies in between, defined by being at least able to switch between sun and dark forever. Fig. 2 summarizes all these regions in the form of a decision tree, where one can identify the region the system is in by answering a small number of questions. That our partitioning is indeed complete and can be given a suitable and unambiguous mathematical form for all kinds of systems is shown in the next section.

While in Fig. 1, each of the introduced set of system states is just one topologically connected region, in general most of these sets are composed of several disjoint regions, so there may be several shelters, glades, lakes, etc. On a finer level, these may be analysed further by looking at which parts may be reached from which other parts, and this leads to a finer, hierarchical partition into *ports, rapids, harbours, docks,* etc. and to several new types of dilemmas, as shown in Fig. 3.

All of the five types of dilemmas listed in Table 1) can easily occur in the collective "management" or governance of the Earth System by humanity. A glade dilemma may occur if adaptation is seen as preferable to mitigation for welfare reasons but turns out to be a riskier option due to a higher uncertainty of the corresponding climate impacts. A lake dilemma can arise if a great transformation of the global energy system towards a carbon-free economy would temporarily lead to welfare losses in poorer countries. A port dilemma may come from the option of increasing welfare by extending industrial agriculture causing biodiversity loss (decreasing flexibility) due to the related large-scale land-use change. A harbour dilemma could occur in the future when colonization of other planets (increasing flexibility) becomes feasible but extremely costly. Finally, a dock dilemma arises whenever a very promising new technology with some unknown risks and side-effects (such as genetically engineered food production) could be introduced on a planetary scale.

2 Formal framework

We will now put all of the above on thorough mathematical footing. Let us assume a *manageable dynamical system with desirable states,* given by the following components:

(i) A dynamical system with a *state space* X, a *default dynamics* represented by a family of *default trajectories* $\tau_x(t)$, and some basic *topology* on X (e.g., the Euclidean topology, see Appendix A1 for more detail).

(ii) A notion of *desirable states* represented by an open set $X^+ \subseteq X$, called the *sunny region,* whose complement $X^- = X - X^+$ we call the *dark.*

(iii) A notion of *management options* represented by a family $\mathcal{M}_x$ of *admissible trajectories* μ for each $x \in X$.

We assume that one can switch immediately to any trajectory $\mu \in \mathcal{M}_x$ whenever in state x. We say the system *floats* when it follows a default trajectory, and that we may *row* the system along any other admissible trajectory.

Note that although, formally, we consider deterministic autonomous systems only, non-deterministic systems can be incorporated by considering probability distributions as states, time-delay systems can be treated similarly, and externally driven or otherwise explicitly time-dependent systems can be covered by including time t as a variable with $\dot{t} = 1$ into the state vector. Also, if management involves some form of inertia, e.g., if not the propelling vector v of a boat but only its acceleration $\dot{v}$ can be changed discontinuously, the proper way to model this in our framework would be to treat v as part of the state.

2.1 Qualitative distinction of regions w.r.t. sustainable manageability of desirability

The main idea of the coarsest of our classifications of states is to first identify (i) a *safe* region where management is unnecessary, called the *shelters* S, and (ii) a less safe but larger *manageable region* M where one can permanently avoid the dark at least by management. Then we classify all states with regard to whether and how X^+, S, and M can be stably reached from the current state by management. For each state, we ask: (iii) Can S be stably reached, and if so, can the dark be avoided on the way? (iv) If not, can M be stably reached? (v) If not, can we stably reach X^+ over and over again, or at least once again? We will see that these criteria lead to a partition of state space into a "cascade" consisting of five main regions, *upstream* U, *downstream* D, *eddies* E, *abysses* Υ, and *trenches* Θ. Each of these will then be split up further into sets such as *glades* G, *lakes* L, and *backwaters* W, etc., by asking further qualitative questions. In choosing these figurative terms, we try to avoid a too technically-sounding language and rather extend the useful and common metaphor of "flows" and "basins" in a natural way without trying to match their common-language meanings too accurately.

To acknowledge the fact that all real-world dynamics and management will be subject to at least infinitesimal noise and errors, we base the formal definition of these state space regions on certain notions of *invariant open kernel, sustainability,* and *stable reachability,* whose symbolic mathematical definitions and algebraic properties are detailed in Appendix A2.

2.2 Shelters, manageable region, upstream & downstream

The *invariant open kernel* of a set $A \subseteq X$, denoted $A^{\iota o}$, is the largest open subset of A that contains the default trajectories of all its own points. The *shelters* are the invariant open kernel of the sunny region,

$$S = (X^+)^{\iota o}. \tag{1}$$

S contains all sunny states whose default trajectories stay in the sunny region X^+ forever without any management even when infinitesimal (or small enough) perturbations occur. In other words, when inside S, one *will* "stably" stay in X^+ *by default.*

We call an open set $A \in \mathcal{T}$ *sustainable* (in the basic sense of the word, simply meaning that it can be sustained) iff it contains an admissible trajectory for each of its points. The *sustainable kernel* of a set $A \subseteq X$, denoted $A^{\mathcal{S}}$, is the largest sustainable open subset of A. We call the sustainable kernel of the sunny region the *manageable region:*

$$M = (X^+)^{\mathcal{S}} \supseteq S. \tag{2}$$

In other words, when inside M, one *can* stably stay in X^+ *by management.*

In Appendix A2, we introduce a suitable notion of stable reachability to overcome two problems with the classical notion of (plain) reachability known from control theory. For now, let us assume we know what we mean when saying that a state y or a set $Y \subseteq X$ is *stably reachable* from some state x *through* some set $A \subseteq X$, denoted $x \rightsquigarrow_A y$ or $x \rightsquigarrow_A Y$. Using this notion of stable reachability for the choice $A = X$ (other choices of A will be used in the next section), we can now define the *upstream* U as the set of states from where the shelters S can be stably reached at all. Likewise, the *downstream* D consists of all states from which the manageable region M but not the shelters can be stably reached:

$$U = (\rightsquigarrow_X S) \supseteq S, \tag{3}$$

$$D = (\rightsquigarrow_X M) - (\rightsquigarrow_X S) = (\rightsquigarrow_X M) - U \supseteq M - U. \tag{4}$$

2.3 Trenches, abysses, eddies & the main cascade

On the other, dark end of what we will call the main cascade, we first define the *trenches* Θ as that region in the dark from which one cannot stably reach the sunny region even once,

$$\Theta = X - (\rightsquigarrow_X X^+) \tag{5}$$

(this concept approximately corresponds to the "catastrophe domains" of Schellnhuber (1998)).

Now we turn to the region from where one cannot avoid ending up in the trenches. We define the *abysses* Υ as the closure of this region, minus the trenches:

$$\Upsilon = \overline{\{x \in X \,|\, \forall \mu \in \mathcal{M}_x \, \exists t \geqslant 0 : \mu(t) \in \Theta\}} - \Theta. \tag{6}$$

The closure is taken since already an infinitesimally small perturbation from a point in this closure can make the trenches unavoidable.

Finally, the *eddies* E are the remainder of X, i.e., the part from where the manageable region cannot be stably reached but the trenches can be avoided:

$$\begin{aligned} E &= X - U - D - \Upsilon - \Theta \\ &= (X - (\rightsquigarrow_X M)) \cap (X - (\Upsilon + \Theta)). \end{aligned} \tag{7}$$

Thus, when in the eddies, even though one can reach the sunny part over and over again, one cannot stay there forever but has to visit the dark repeatedly.

A connected component of Θ, Υ, or E will be called an individual *trench, abyss,* or *eddy,* and the latter two typically have sunny and dark parts.

The system $\mathcal{C} = \{U, D, E, \Upsilon, \Theta\}$ is a partition of X which we call the *main cascade* because of the following mutual reachability restrictions:

$$U \not\leftsquigarrow_X D \not\leftsquigarrow_X E \not\leftsquigarrow_X \Upsilon \not\leftsquigarrow_X \Theta. \tag{8}$$

In other words, one might at best be able to go in the "downstream" direction by default or by management, from upstream to downstream to the eddies to the abysses to the trenches, but not in the other, "upstream" direction (see also Fig. 2).

2.4 The glades and lake dilemmas, backwaters, and the manageable partition

Some of the states in the manageable region M may be in $U = (\rightsquigarrow_X S)$ but not in $(\rightsquigarrow_{X^+} S)$. This motivates the definition of two subsets of M via the relation of *sunny stable reachability*, $\rightsquigarrow_{X^+}$, namely (i) the *glades* G, from where the shelters can be stably reached through the sun, and (ii) the *lakes* L, from where the shelters can be stably reached only through the dark:

$$G = (\rightsquigarrow_{X^+} S) - S, \tag{9}$$

$$L = M \cap U - (\rightsquigarrow_{X^+} S) = M \cap U - S - G. \tag{10}$$

Glades and lakes are two particularly interesting types of regions since in both one has a qualitative decision problem. The *glade dilemma* occurs if a glade is for some reason more desirable than its shelter, since then one has to decide whether to stay in the more desirable but unsafe glade or row to the less desirable but safe shelter. The *lake dilemma* exists in every lake: shall one stay in the sun by rowing over and over again, but risking to float into the dark if the paddle breaks, or shall one move into a shelter, accepting a temporary passage through the dark, to be able to recline in safety eventually? In other words, the lake dilemma is a choice between uninterrupted desirability and eventual safety. Below we will encounter more qualitative dilemmas of this and other types.

While $\{S, G, L\}$ is a partition of $M \cap U$, also the downstream D may contain a manageable part, the *backwaters* W. This is the region where one may stay in the sun forever by rowing over and over again, but where one may not stably reach the shelters at all, not even through the dark:

$$W = M \cap D = M - U. \tag{11}$$

This completes the *manageable partition*

$$M = S + G + L + W. \tag{12}$$

Also, both U and D may contain points outside M, which we call the *dark upstream/downstream,*

$$U^- = U \cap X^-, \quad D^- = D \cap X^-, \tag{13}$$

and the *remaining sunny upstream/downstream,*

$$U^{(+)} = (U \cap X^+) - M, \quad D^{(+)} = (D \cap X^+) - M, \tag{14}$$

leading to the *upstream* and *downstream* partitions

$$U = S + G + L + U^{(+)} + U^-, \quad D = W + D^{(+)} + D^-. \tag{15}$$

Finally, one can divide the eddies and abysses into sunny and dark parts:

$$E^\pm = E \cap X^\pm, \quad \Upsilon^\pm = \Upsilon \cap X^\pm. \tag{16}$$

All the sets introduced so far are summarized in Fig. 2 in the form of a decision tree that allows for a fast classification of individual states.

2.5 Finer distinction of regions w.r.t. mutual reachability of different types

In addition to the glade and lake dilemmas introduced above, there exist at least three further types of qualitative decision problems, all related to the question of which parts or subregions of the above introduced regions may be stably reached from which other parts, and whether corresponding transition pathways exist that do not leave the shelters or at least the sunny region, or only through the dark. In order to study these questions, we introduce three additional, successively finer partitions derived from the reachability relations $\rightsquigarrow_X$ (stable reachability) and $\rightsquigarrow_{X^+}$ (stable reachability through the sun) that we used already above, and from the even more restrictive relation $\rightsquigarrow_S$ (stable reachability through the shelters).

2.5.1 The ports and rapids partition & network, and the port dilemma

While from each state in U, one can stably reach some part of S, one cannot in general navigate freely inside S or U or any other member of the main cascade $\mathcal{C}$. Let us call a maximal region in which one can navigate freely a *port* (see Appendix A3 for more thorough formal definitions and proofs of the claimed properties). Each port is completely contained in one of the sets $U, D, E, \Upsilon^-, \Theta$, and none can intersect Υ^+, so the notion of ports fits well into the hierarchy of regions that began with the main cascade and the manageable partition. But there are also *transitional* states not belonging to any port since one cannot return to them. So, to extend the system of all ports into a partition of all of X, we also have to classify these non-port states, and we do so by asking which ports they can reach and from which ports they can be reached. States that are equivalent in this sense form what we call a *rapid*. It turns out that U and D are then partitioned into ports and rapids, and so is each individual eddy, abyss, and trench. The reachability relations between ports and rapids form a directed network that concisely summarizes the overall structure of all management options.

Fig. 1 shows the very simple case of a linear network: the whole upstream is one port, the sunny downstream and the adjacent fast-moving part of the dark downstream form a rapid, the backwater and the slow-moving part of the dark downstream form another port, the waterfall is another rapid, the eddy is a port again, and the abyss and the trench are rapids. In the examples below, we will however see that much more complex ports and rapids networks may occur in models, and one can prove that any acyclic graph may occur as the ports and rapids networks of some system.

The ports and rapids partition is helpful in the discussion of a certain type of dilemma that results from two different objectives which may not be easily balanced: (i) the objective of being in or reaching a state with high *intrinsic desirability,* e.g., as measured by some qualitative preference relation finer than the mere distinction between "desirable" and "undesirable", or even by some quantitative evaluation such as a welfare function; and (ii) the objective of retaining an amount of *flexibility* as large as possible by being in or reaching a state from which a large part of state space is reachable. Flexibility may be important in particular in situations in which there is some uncertainty about future management options and/or future preferences (Kreps, 1979). We call this a *port dilemma.*

2.5.2 The harbours and channels partition & network, and the harbour dilemma

Since it does not take into account the definition of the desirable region X^+ at all, ports and rapids are not directly compatible to the regions from the manageable partition $\mathcal{M}$ since their members may overlap in complex ways. However, we can construct a very similar but finer partition based on stable reachability through the sun ($\rightsquigarrow_{X^+}$) instead of (plain) stable reachability, restricted to the sunny region, and the result turns out to be compatible with $\mathcal{M}$.

A maximal region in which one can freely navigate without leaving the sun is called a *harbour.* A region of states that do not belong to any harbour but from which the same harbours can be reached through the sun and which can be reached from the same harbours through the sun is called a *channel*. Since each harbour or channel lies completely in a port or a rapid, the harbours and channels form a finer partition than the ports and rapids and form a finer layer of the reachability network in which the links represent reachability through the sun instead of mere reachability.

The harbours-and-channels partition allows one to identify decision problems involving (i) the objective of *staying* in a desirable state and (ii) the objective of eventually *reaching* a state with higher desirability or flexibility, which is called a *harbour dilemma* here.

2.5.3 The docks and fairways partition & network, and the dock dilemma

Note that although the harbours-and-channels partition is finer than that into ports and rapids, there is still one important region that can have nontrivial overlaps with harbours and channels, namely the shelters S. In order to complete our hierarchy of partitions and networks of regions, we therefore introduce a third and finest partition and network level, restricted to S, based on the notion of *stable reachability through the shelters,* $\rightsquigarrow_S$.

In complete analogy to the above, a maximal region of states that are mutually reachable through S is called a *dock,* and the non-dock states in S are classified into so-called *fairways* with regard to their reachability of these docks. Again, each dock or fairway lies completely in a harbour or channel, and they form a third layer of the reachability network whose links now represent the safest form of reachability, namely through the shelters.

Finally, the docks-and-fairways partition is helpful in the discussion of dilemmas involving (i) the objective of staying in a *safe* state (i.e., in the shelters) and (ii) the objective of eventually reaching a state with higher desirability or flexibility. We call this a *dock dilemma.*

2.6 Summary of the introduced hierarchy of partitions and networks

Summarizing, we have now a hierarchy of ever-finer partitions of the system's state space at our hands. We began with the main cascade $\mathcal{C} = \{U, D, E, \Upsilon, \Theta\}$, its refinement into the partition $\{S, G, L, U^{(+)}, U^-, W, D^{(+)}, D^-, E^+, E^-, \Upsilon^+, \Upsilon^-, \Theta\}$ (see Fig. 2), and the further refinement by topological connectedness into individual shelters, glades, lakes, backwaters, eddies, abysses, and trenches. These partitions represent the qualitative differences in stable reachability of the shelters or the manageable set, thus allow for a first classification of states w.r.t. the possibilities of sustainable management, and may reveal decision problems of the type of glade or lake dilemma which will occur in many of the examples below, where one has to choose between higher safety and higher desirability or flexibility or between uninterrupted desirability and eventual safety.

A different refinement of $\mathcal{C}$ into the ports-and-rapids network is still based on stable reachability alone but contains other details suitable for the identification and discussion of possible port dilemmas that involve a choice between higher desirability and higher flexibility. Inside the desirable region X^+, this partition can be refined into the harbours-and-channels network suitable for the discussion of harbour dilemmas that involve a choice between uninterrupted desirability and eventually higher desirability or flexibility, and further into the docks-and-fairways network suitable for the discussion of dock dilemmas that involve a choice between uninterrupted safety and eventually higher desirability or flexibility (Table 1).

These three networks may also be interpreted as a three-level "network of networks" with nodes representing state space regions of different quality and size. A network-theoretic analysis of it using methods such as the node-weighted measures of Heitzig et al. (2012) may especially be interesting in the context of varying system parameters and bifurcations such as those in Fig. B2, but is beyond the scope of this article.

3 Examples

In this section, we will apply the introduced framework to several illustrative examples from natural and co-evolutionary Earth System modeling, ecology, socio-economics, and classical mechanics. The examples have been chosen not for their realism but for their simplicity, to show the broad scope of potential applicability of our concepts, and the relevance of the identified types of decision dilemmas in both the natural and socio-economic components of the Earth System.

3.1 Carbon cycle & planetary boundaries

Our first example is from natural Earth System modeling and illustrates which of the above-introduced regions occur most often for systems that possess only a single, globally stable, and desirable attractor.

Anderies et al. (2013) proposed a conceptual model of the global carbon cycle capturing its main features while keeping the model sufficiently low-dimensional to be able to discuss the planetary boundaries concept with it. We use their model for pre-industrial times, which has three dynamical variables c_m, c_t and $c_a = 1 - c_m - c_t$ representing the maritime, terrestrial, and atmospheric shares of the fixed global carbon stock. The dynamics is of the form

$$\dot{c}_m = a_m(c_a - \beta c_m), \quad \dot{c}_t = f(c_a, c_t) - \alpha c_t,$$

where a_m, β are diffusion parameters, f is a function representing photosynthesis and respiration, and α governs the human offtake rate from the terrestrial carbon stock. See Anderies et al. (2013) for details and parameter values.

Since the parameter α can be considered the natural human management option for this system, we assume the default flow has a value of $\alpha = \alpha_+ = 0.5$, while management can reduce it by half to $\alpha = \alpha_- = 0.25$, which results in the trajectories shown in Fig. 4. Both have a unique stable fixed point in the interior of the state space which is globally attractive for all states with $c_t > 0$.

In order to roughly represent the planetary boundaries relating to climate change, biosphere integrity, and ocean acidification (Rockström et al., 2009b; Steffen et al., 2015), we require a "sunny" state to have sufficiently low atmospheric carbon, at least a minimum value of terrestrial carbon, and not too large maritime carbon, leading to a dark region of the shape shown in Fig. 4 in grey. If, as shown, the unmanaged fixed point is sunny, one obtains a purely upstream situation with a shelter surrounding the fixed point, a glade, and a remaining sunny upstream $U^{(+)}$ as shown in the figure. For our (quite arbitrarily) chosen parameter values, a trajectory starting in the sunny upstream is likely to first cross the climate boundary and then the biosphere boundary before getting back into the sunny region, whereas it seems quite unlikely to cross the acidification boundary.

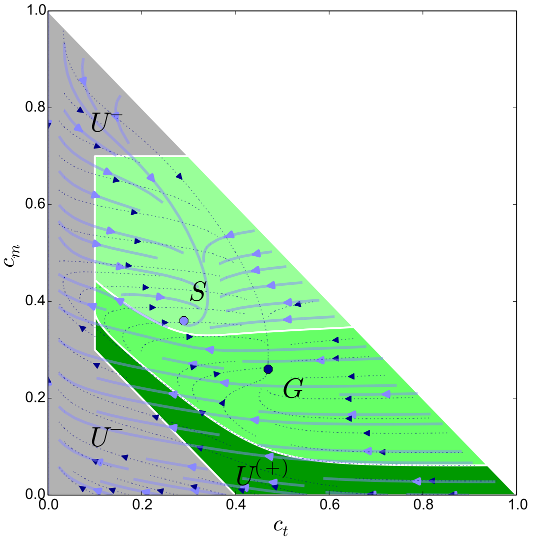

Figure 4. Phase portrait of the pre-industrial carbon cycle model of Anderies et al. (2013). Arrows indicate default/unmanaged dynamics (pale blue) and alternative/managed dynamics (dotted dark blue) from reducing the human offtake rate by half. Filled dots: corresponding stable fixed points. Grey area: undesired region defined by (i) upper bounds for maritime carbon c_m (white horizontal line, representing a planetary boundary related to ocean acidification) and atmospheric carbon $1 - c_t - c_m$ (white diagonal line, related to a climate change boundary) and a lower bound for terrestrial carbon c_t (white vertical line, representing an ecosystem services planetary boundary). Coloured areas and labels: derived state space partition (see text), colors as defined in Fig. 2: a shelter S around the globally stable fixed point of the default dynamics, a glade G from where S can be reached by management without violating the bounds, and a remaining sunny upstream $U^{(+)}$ from where one cannot avoid violating the bounds temporarily.

In this example, all non-upstream regions are empty, and so is the lake region, hence no lake dilemma occurs. On the other hand, if one considers a higher c_t to be preferable, we get an example of the glade dilemma since the managed fixed point in the less safe glade has higher c_t than the unmanaged fixed point in the safer shelter. Note that this is neither a port,

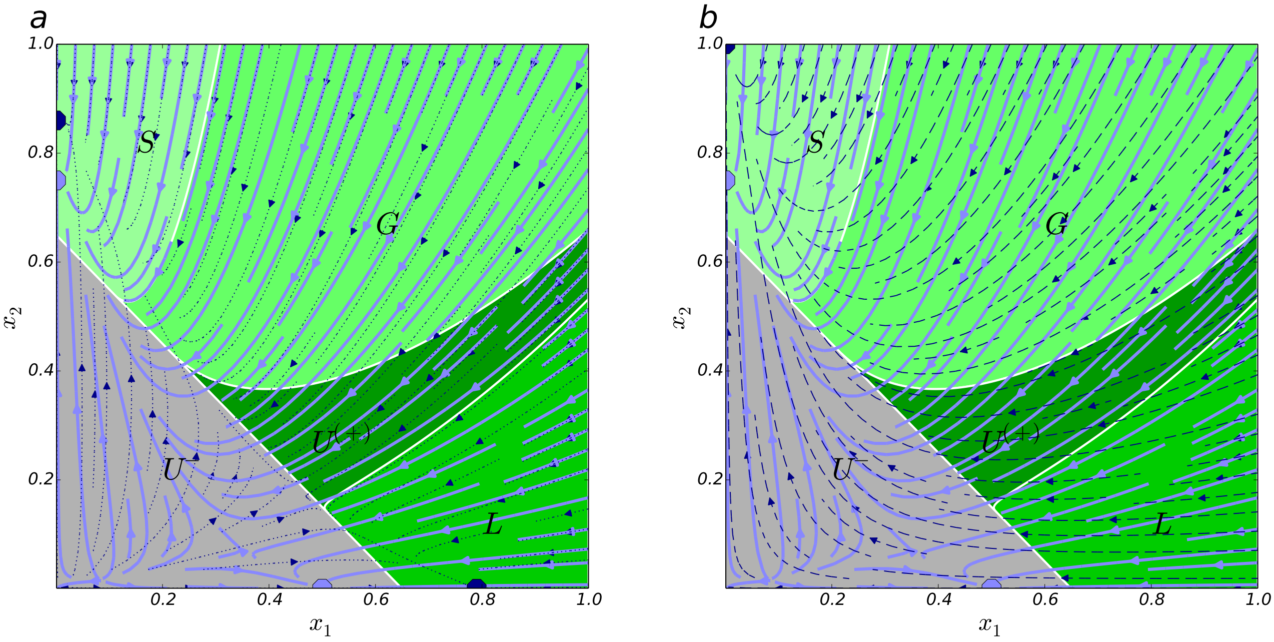

Figure 5. Competing plant types example, showing all upstream regions and illustrating the lake dilemma. A bistable system of two competing plant types with two simultaneous management options (depicted in separate plots only for discernability). Management by a general harvesting quota (dotted arrows shown left) can ensure desirable long-term harvests of the less productive type x_1 (*lake* L). Management by temporary protection of the more productive type x_2 (dashed arrows shown right) can cause a transition to the desirable fixed point (in the *shelter* S), but only through the undesired region of low harvests (gray region). The state space partition boundaries resulting from both options together (white curves) and a desirable minimum harvest boundary (white diagonal) follow some admissible trajectory at each point.

harbour, or dock dilemma since both points are in the same port and harbour and only the unmanaged one is in a dock.

If, instead, we had chosen the minimum value for c_t to be larger than the unmanaged equilibrium value, the shelter would be empty and the whole situation would change from upstream-only to either a downstream-only or an abyss-and-trench situation. This type of *topological bifurcation* will be studied in Example 3.4. In the next example, we will see a lake dilemma instead of a glade dilemma.

3.2 Competing plant types & multistability

The second example, from ecology, demonstrates how the lake dilemma may occur in a multistable system with a sunny and a dark attractor.

In this fictitious example, two plant types $1,2$ compete for some fixed patch of land, modify the soil, and are harvested. Their growth follows a logistic-type dynamics, with land cover proportions $x_{1,2} \in [0,1]$ following the equations

$$\dot{x}_1 = x_1(K_1(x_{1,2}) - x_1) - h_1 x_1,$$
$$\dot{x}_2 = r x_2(K_2(x_{1,2}) - x_2) - h_2 x_2.$$

In this, $r > 1$ is a constant productivity quotient, $h_{1,2}$ are the harvest rates, and the two dynamic capacities $K_1(x_{1,2}) = \sqrt{x}_1(1-x_2) \leqslant 1$ and $K_2(x_{1,2}) = \sqrt{x}_2(1-x_1) \leqslant 1$ represent the fact that each type modifies the soil quickly to its own benefit but to the other type's disadvantage (see Supplement 1 for a discussion of the model design).

For our illustration, we assume that on the default trajectories, both harvest rates $h_{1,2}$ equal some rather high value h_+, leading to low equilibrium harvests. We assume management can repeatedly choose between this default and two types of alternative trajectories. Type 1 has a lower value for both harvest rates, $h_{1,2} = h_- < h_+$, representing management by restricting harvests politically in order to yield higher long-term harvests, but without aiming to change the plant mix, as depicted in Fig. 5 (left). Type 2 management option has harvest rates $h_2 = 0$ and $h_1 = 2h_+$, representing management by temporarily protecting type 2 in order to change the plant mix to the higher productivity plant; we assume that this moratorium results in more intense harvesting of type 1, as depicted in Fig. 5 (right). We assume that both options exist simultaneously at all times (the separate plots of Fig. 5 are only for better discernability of the trajectories). We set the desirable region to where $x_1 + x_2 > \ell$ for some $\ell > 0$ in order to ensure some minimum harvests.

For the choice $r = 2$, $h_+ = 0.2$, $h_- = 0.1$, $\ell = 0.65$ of the figure, the desirable high productivity stable fixed point of the default dynamics at $\approx (0, 0.79)$ is in the sunny region and is thus contained in a shelter S. The latter is delimited by the default trajectory that meets the boundary to the undesired region tangentially. S can be stably reached from all states with $x_2 > 0$, hence the upstream is $U = \{(x_1,x_2)|x_2 > 0\}$. The border of the glade G next to S can be found by backtracking the "widest" admissible trajectory that meets the boundary to the undesired region tangentially; this turns out to be a type-2 management trajectory as seen in Fig. 5 (right). This shows how the boundaries of regions may often be found by identifying tangential or otherwise significant points and backtracking the default and alternative trajectories leading to them.

The lower productivity stable fixed point of the default dynamics (with $h_{1,2} = h_+$) at $\approx (0.52, 0)$ is undesired for this choice of X^+. From it one can not only navigate to S but can also (and faster) get to the higher productivity stable fixed point of the first type of *managed* dynamics with $h_{1,2} = h_-$, at $\approx (0, 0.79)$, and stay there as long as management holds. Hence the region around $(0, 0.79)$ is part of the manageable region M. The exact boundary of this region (which soon turns out to be a lake, L) is the "widest" admissible trajectory that meets the boundary to the undesired region tangentially; in this case, this trajectory turns out to be a type-1 management trajectory as seen in Fig. 5 (left). To get from this type 1-dominated region to the type 2-dominated shelter S via the other management option of protecting type 2, one has to cross the undesired middle region in which both types coexist at a low level due to soil conditions that are suboptimal for both types. Hence the region around $(0, 0.79)$ is a lake. The associated lake dilemma is similar to a glade dilemma in that staying in a lake is unsafe as in a glade, but it differs in the reason why one may want to stay there: While staying in a glade may be attractive simply because the glade may be more desirable than the shelter in some quantitative sense, staying in a lake may seem attractive since that avoids having to pass through the dark to reach safety.

This form of the lake dilemma can also occur in other multistable systems when one of the attractors is in the dark but sufficiently close to the sunny region so that constant management can sustain the system in a sunny place near that attractor, and when other management options may push the system towards another, sunny attractor after crossing the dark.

Note that in this example, the lake dilemma falls together with a port dilemma since after leaving the lake for the shelter, one cannot return. If we choose a slightly larger sunny region by lowering ℓ to $\ell = 0.45$, the unmanaged fixed point with $y = 0$ gets into X^+ and the former lake around it now becomes a second shelter, which might be called a *shelter/lake transition.* But from this shelter the other, more desirable shelter can still only be reached through the dark. Since the two shelters correspond to two harbours in the reachability network, this means the former lake dilemma has been converted into a harbour dilemma.

The example also shows that the more management options exist, the less trivial it is to find the boundaries between regions even in two-dimensional systems. For higher dimensions, one will usually have to rely on specialized numerical algorithms such as the Viability Kernel Algorithm of Frankowska and Quincampoix (1990) from viability theory.

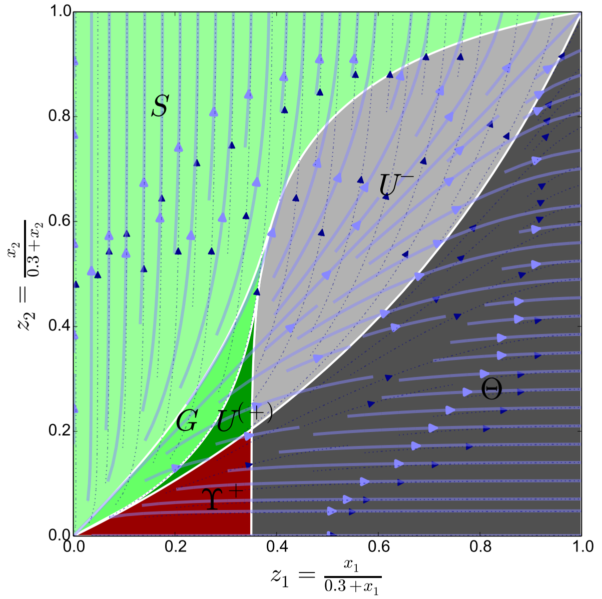

Figure 6. Substitution of a dirty technology. Coevolution of the cumulative production of a dirty technology x_1 and a clean one (x_2) without (pale blue curves) and with (dotted dark blue curves) a subsidy for the clean technology. Undesired region with too high future usage of the dirty technology colored in grey. Knowledge stocks $x_{1,2}$ were transformed to $z_{1,2} = x_{1,2}/(0.3 + x_{1,2})$ in order to capture their divergence to $+\infty$.

3.3 Substitution of a dirty technology

Our third example concerns a purely socio-economic part of the Earth System that bears some similarity to the preceding example but features regions from both ends of the main cascade: upstream and abyss/trench, without having the intermediate regions of downstream and eddies.

Instead of plants, in this example a certain produced good (e.g., electric energy) comes in two types which are economically perfectly substitutable but whose production processes use two different technologies, one "dirty" and one "clean" (e.g., conventional and renewable energy). The production costs C_1, C_2 are convex functions of production output per time y_i and decrease over time via a learning-by-doing dynamics that is similar to Wright's law (Nagy et al., 2013):

$$C_i(y_i) = \gamma_i y_i^{1+\sigma_i}/(1+\sigma_i)x_i^{\alpha_i}.$$

In this, x_i is cumulative past production (with $\dot{x}_i = y_i$), γ_i are cost factors, $\sigma_i > 0$ are convexity parameters, and $\alpha_i > 0$ are learning exponents. We assume that demand D depends linearly on price, $D(p) = D_0 - \delta p$, $\delta > 0$, that demand equals production, $D = y_1 + y_2$ (“market clearance”), and that price equals marginal costs, $p = \partial C / \partial y_i = \gamma_i y_i^{\sigma_i} / x_i^{\alpha_i}$, due to perfect competition among producers. One can then uniquely solve for the produced amounts y_i, getting some formula $y_i = f_i(x_1, x_2)$. This results in a two-dimensional dynamical system with state variables x_1, x_2 and equations

$$\dot{x}_i = f_i(x_1, x_2).$$

Let us put $D_0 = 1$, $\delta = 1$, $\sigma_i \equiv 1/5$, $\alpha_i \equiv 1/2$, and assume that the default dynamics has $\gamma_i \equiv 1$, so that the long-term default behaviour is $p(t) \to 0$, $D(t) \to 1$. If the dirty technology 1 is the traditional one, so that $x_1(0) > x_2(0)$, we have $x_1(t) \to \infty$, $x_2(t) \to \hat{x}_2 < \infty$, $y_1(t) \to 1$, and $y_2(t) \to 0$, i.e., usage of the clean technology 2 will die out. If instead $x_1(0) < x_2(0)$, technology 1 will die out. Hence the system is bistable as in the plant example, but with attractors at infinity. To depict the diverging behaviour, we used the transformation $z_i = x_i/(1 + x_i)$ in Fig. 6.

The main dynamical difference to the plant example is however not the diverging behaviour, but has to do with the choice of management options. While in the plant example, the choice of management options led to an upstream-only situation in which the more desirable fixed point could be reached from everywhere, in this example we will get regions from which the desirable fixed point cannot be reached and which are thus non-upstream. We consider the management option of lowering γ_2 to a value of, say, $1/2$ by subsidising the clean technology 2 to induce a technological change (Jaffe et al., 2002; Kalkuhl et al., 2012). This leads to the alternative dynamics depicted in Fig. 6, showing that for some initial states with $x_1 > x_2$ one can now get $x_2(t) \to \infty$ and $y_1(t) \to 0$. The goal of keeping the usage of the dirty technology below some limit, $y_1 < \ell < 1$, corresponds to a desirable region in terms of x_1, x_2, whose border can be computed as $x_2 = x_1(1/\ell - 1 - 1/\ell^{4/5}\sqrt{x_1})^{2/5}$, see Fig. 6. That goal is automatically fulfilled in the top-left shelter region, can also be sustained by management (subsidies) in the glade region below it, and can at least be reached eventually from the remaining sunny upstream $U^{(+)}$ below the glade and from the dark upstream U^- which is delimited by the management trajectory that meets the upper right corner.

But from below the latter trajectory, the shelter cannot be reached. In other words, when in U^-, one has to act fast in order not to loose the option of reaching S. From the dark part denoted Θ, not even the sunny region be reached, hence that region is a trench, while the sunny part to its left is the abyss leading to that trench. There are no intermediate regions (downstream or eddies) between upstream and abyss in this example.

3.4 Combined population and resource dynamics

Our fourth example models the coevolution (in the sense of joint time evolution) of a natural Earth System component coupled with a socio-economic Earth System component and

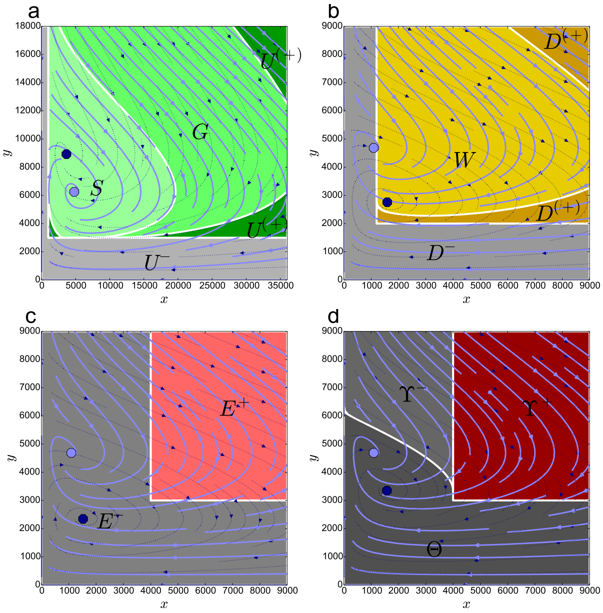

Figure 7. Combined population and resource dynamics. Coevolution of a population x and a resource stock y. In all cases, $\phi = 4, r = 0.04$. When the globally stable fixed point of the default dynamics (pale blue) falls into X^+, only upstream regions occur (top-left, $\gamma_0 = 4 \cdot 10^{-6} > \gamma_1 = 2.8 \cdot 10^{-6}, \delta = -0.1, \kappa = 12000, x_{\min} = 1000, y_{\min} = 3000$). When it falls into X^- instead, but the stable fixed point of the alternative management trajectory (dotted dark blue) is in X^+, then only downstream regions occur (top-right, $\gamma_0 = 8 \cdot 10^{-6} < \gamma_1 = 13.6 \cdot 10^{-6}, \delta = -0.15, \kappa = 6000, x_{\min} = 1200, y_{\min} = 2000$). Otherwise (bottom, $\gamma_0 = 8 \cdot 10^{-6} < \gamma_1, \delta = -0.15, \kappa = 6000, x_{\min} = 4000, y_{\min} = 3000$), the analysis depends on whether one can repeatedly reach X^+ by switching between default and alternative trajectories: For $\gamma_1 = 16 \cdot 10^{-6}$ (bottom-left), only eddies occur, while for $\gamma_1 = 11.2 \cdot 10^{-6}$ (bottom-right), only abysses and trenches occur.

shows how different parameters may qualitatively move the resulting state space topology through the whole main cascade, from an upstream-only situation via downstream-only and eddies-only to an abyss-and-trench situation.

The model was used in Brander and Taylor (1998) to explain the rise and fall of the native civilization on Rapa Nui (Easter Island) before western contact, but it may also be interpreted as a conceptual model of global population-vegetation interactions. It is derived from simple economic principles and leads to a a modified Lotka-Volterra model with a finite resource. The human population x is preying on the island's forest stock y which itself follows a logistic growth dynamics:

$$\dot{x} = \delta x + \phi \gamma x y, \quad \dot{y} = r y (1 - y/\kappa) - \gamma x y$$

for some parameters $\gamma, \delta, \kappa, \phi, r$ representing growth and harvest rates and the stock's capacity.

We assume management will either reduce the default harvest rate γ_0 to some smaller value $\gamma_1 < \gamma_0$ to avoid over-

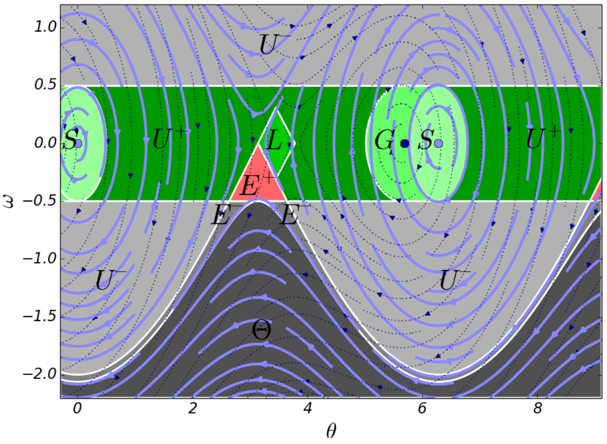

Figure 8. Gravity pendulum fun-ride with management by one-sided acceleration and undesirable fast rotations. The 2π-periodic coordinate θ is the pendulum's inclination angle. If its angular velocity ω exceeds $\pm\ell$, people get sick (grey region). Since staying in L (balancing almost upright) or G (balancing somewhat inclined) is more exciting than in S (resting downward), we have both a glade and a lake dilemma.

exploitation of the resource, or increase it to a larger value $\gamma_1 > \gamma_0$ to avoid famine. Our choice of the sunny region relies on two principles. The absolute population should not drop below a threshold $x_{\min}$ and the relative decline in population under the default dynamics, $-\dot{x}/x$, should not exceed a value of ℓ. Hence $X^+ = \{x > x_{\min}$ and $y > y_{\min} = \max(0, -(\ell+\delta)/\phi\gamma_0)\}$.

The resulting state space partition is depicted in Fig. 7 for $\phi = 4, r = 0.04$ and different choices of $\gamma_0, \gamma_1, \delta, \kappa, x_{\min}, y_{\min}$. One either gets an upstream-only situation, a downstream-only one, an eddy-only one, or an abyss-and-trench situation, depending on whether the unmanaged and managed fixed points belong to the desired or undesired region. In Appendix B2, these kinds of transitions are more formally interpreted as bifurcations.

An interesting case occurs when the whole state space is a single eddy as in Fig. 7 (bottom-left): One can then repeatedly visit the sunny region by suitably switching between a low default harvest rate and a managed higher harvest rate, but one cannot avoid getting back into the undesired region of a low or fast declining population. An "optimal" management strategy would then lead to slowly but strongly oscillating behaviour.

3.5 Gravity pendulum fun-ride

While in the above examples typically only some of the possible regions were non-empty for each parameter combination, the following example from classical mechanics displays a rich diversity of state space regions that coexist at a single choice of parameter values. Despite an extremely simple dynamics, it features both a glade and a lake dilemma, an eddy, and a trench at the same time.

In the model, people sit in a fun ride resembling a gravity pendulum with angle θ and angular velocity ω and default dynamics given by

$$\dot{\theta} = \omega, \quad \dot{\omega} = -\sin\theta.$$

An optional additional clockwise acceleration of the pendulum of magnitude $a > 0$ ("management") leads to alternative admissible trajectories on which for some time interval(s) one has $\dot{\omega} = -\sin\theta - a$. The sunny region is where $|\omega| < \ell$, for some $\ell > 0$ representing a safety speed limit above which people might get sick.

The unique shelter S is delimited by the default trajectory leading through the points $\theta = 2k\pi$, $\omega = \pm\ell$ that surrounds the stable resting state of $\theta = \omega = 0$, see Fig. 8. If a state lies on a default trajectory that has $\omega > 0$ (counterclockwise pendulum motion) at least some of the time, then there is an admissible trajectory from it leading into the shelter, generated by the management strategy of "braking" whenever $\omega > 0$. Hence the upstream U equals the region strictly above the default trajectory with $\omega < 0$ that connects the unstable saddle point at $\theta = (2k+1)\pi$, $\omega = 0$ (pendulum balancing upright) with itself.

Just left of the shelter is the unique glade G. Depending on the parameter values, the stable fixed point of the managed dynamics (hanging pendulum inclined by constant acceleration) may either belong to the shelter or to the glade. In the latter case (Fig. 8), we have a glade dilemma since the inclined position is preferred to the resting position by the riders but is unsafe since when the engine breaks, people will get sick.

An even more exciting position is close to the upright balancing saddle point, at θ slightly larger than $(2k+1)\pi$ and $\omega \ll 1$, where there is an admissible trajectory that stays close to there (by braking repeatedly for short intervals while staying almost upright), so that this point is in the manageable region M. This is a typical example of how a region close to a saddle point of the default dynamics may become manageable due to an alternative feasible trajectory that has a slightly *shifted saddle point,* so that in the diamond-shaped region between the two saddle points, one can concatenate unmanaged and managed trajectories into periodic orbits.

However, for choices such as $a = 0.6$ and $\ell = 0.5$ (Fig. 8), there is no admissible trajectory leading from the exciting region with $\theta \approx (2k+1)\pi$, $\omega \approx 0$ into the shelter without entering the region with $|\omega| > \ell$. In that case the diamond-shaped region is a lake and we have a lake dilemma.

Finally, the region below and including the default trajectory that touches the line $\omega = -\ell$ from below is the trenches since one cannot brake in that direction, and the region between the trench and the upstream is the eddies. Downstream and abysses are empty in this example.

3.6 Bifurcations with manageable parameter

This final example system is designed to illustrate the relationship of reachability and bifurcations of a dynamical system that can be managed through a parameter and shows bifurcations of the type typically associated with tipping elements of the Earth System (Schellnhuber, 2009).

It has a two-dimensional state space $X = \{(r, y)\}$, where the "fast" variable $y \in \mathbb{R}$ has a default dynamics

$$\dot{y} = h(y|r) = -(4+r^2)^3 y^3 + (2r^2-1)(4+r^2)y + e^r - 10$$

that cannot be managed directly, and $r \in \mathbb{R}$ is a "slow" variable with (approximately) no default dynamics ($\dot{r} = 0$) which however can be changed by management up to a velocity at most 100 and with arbitrarily large acceleration, leading to admissible trajectories with $\dot{r} \in [-100, 100]$ and $\dot{y} = h(y|r)$. We assume that values of $y \leqslant -1/3$ are undesirable.

If r is instead interpreted as a parameter of the one-dimensional system $\dot{y} = h(y|r)$, the set X can be interpreted as its bifurcation space in which one can plot a bifurcation diagram consisting of the loci of stable (solid lines) and unstable (dotted lines) fixed points, as shown in Fig. 9. As one can see, there are three saddle-node bifurcations at $r_1 \approx -2.2$, $r_2 \approx 1.735$, and $r_3 \approx 4.9$ with monostable parameter regimes $r_1 < r < r_2$ and $r > r_3$, and bistable parameter regimes $r < r_1$ and $r_2 < r < r_3$. Individual and paired saddle-node bifurcations (with often result from fold bifurcations) occur frequently in bistable Earth System components such as the hysteretic Thermohaline Circulation (Stommel, 1961; Rahmstorf et al., 2005), monsoonal soil-vegetation feedbacks (Janssen et al., 2008), or other tipping elements (Schellnhuber, 2009). Hysteresis also occurs on other spatial and temporal scales, e.g. in local hydrology (Beven, 2006) and in long-term glacial climate dynamics (Ganopolski and Rahmstorf, 2001).

The main part of the resulting network of ports and rapids of our example system is depicted in Fig. 10. On its coarsest level, there are two ports, each containing one of the two connected loci of stable/unstable fixed points, and a rapid in between through which one can pass from the left to the right port but not back. If the right port seems more attractive, e.g. because it allows a higher value of y, we have a port dilemma since by leaving the left port for the right one, we loose flexibility in terms of reachable regions.

The right port contains two harbours, similarly connected by a narrow "internal" channel, but also contains another "exit" channel leading from the right harbour to the dark region. Note that on the leftward pointing dashed management trajectory in the middle of the bifurcation diagram, there is a leftmost point from where one can still "turn around" and reach (if only unstably) the right part without entering the dark region; this point is a corner of the right harbour (but not belonging to it, for stability reasons), and below it is a channel leading to another harbour in the bottom-left. Again, if the right harbour seems more attractive, we have a dilemma, this time a harbour dilemma since in order to reach the right harbour from the left one, we have to pass through the dark.

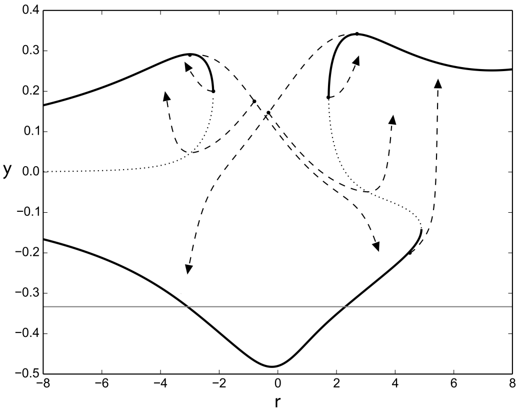

Figure 9. Bifurcations with manageable parameter. Loci of stable (solid black lines) and unstable (dotted lines) fixed points of $\dot{y} = -(4+r^2)^3 y^3 + (2r^2-1)(4+r^2)y + e^r - 10$. Leftmost and rightmost admissible management trajectories (dashed arrows) and their starting points (dots). Border (grey line) between sunny region $y > -1/3$ and the dark. See Fig. 10 for an analysis.

Finally, the right harbour contains two docks again connected by a fairway, plus some more fairways. Again, we get a dilemma if the top-right dock is more attractive than the top-left one: the dock dilemma is that in order to reach the top-right dock from the top-left one, one has to pass through the unsafe middle region and risk ending up in the dark if management breaks down.

4 Discussion and Conclusions

We have presented a formal classification of the possible states of a dynamical system such as the Earth System into regions of state space which differ qualitatively in their safety, the possibilities of reaching a safe state, the possibilities of avoiding undesired states, and in the amount of flexibility for future management.

Based on an assumed main division of the system's states into only two classes, desirable ("sunny") and undesirable ("dark"), we have constructed a hierarchy of partitions of a system's state space, whose member regions we suggested to name by metaphorical names either corresponding to the general image of a boat floating or rowing on a complex water system, such as "upstream", "downstream", "eddy", "abyss", "trench", "lake", and "backwater", or corresponding to the image of a "shelter" surrounded by a "glade". To capture the nature of and relationships between the different regions, we have introduced the notion of stable reachability and the corresponding three-level reachability network of "ports", "harbours", "docks", "rapids", "channels", and "fairways", and illustrated our concepts with conceptual example models from climate science, ecology, coevolutionary

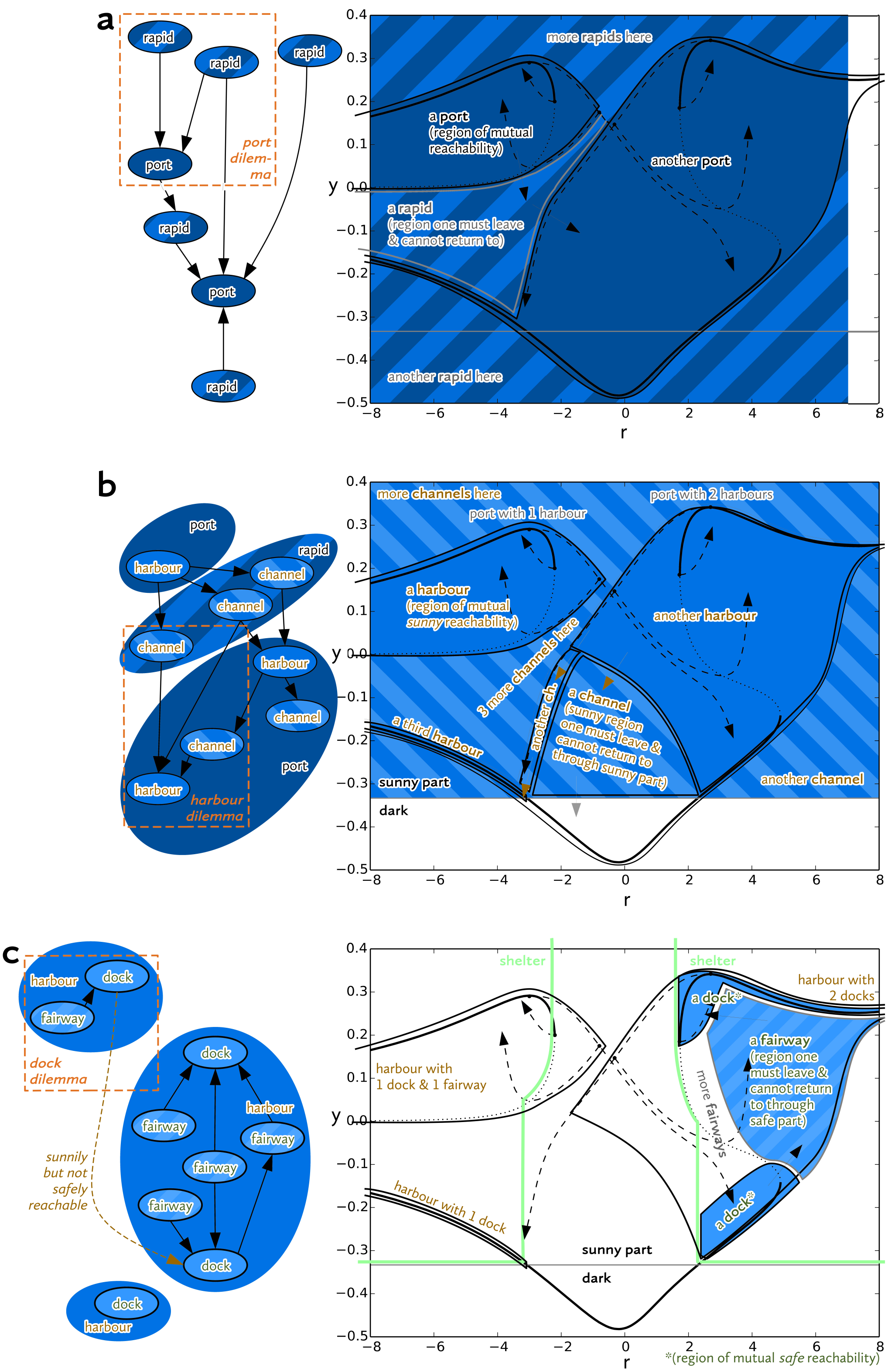

Figure 10. Main part of the three-level reachability network of ports and rapids (top), harbours and channels (middle), and docks and fairways (bottom), and related dilemmas in the bifurcation example. Arrows indicate stable reachability (top), stable reachability through the sun (middle), and stable reachability through the shelters (bottom). Some further arrows between rapids, channels and fairways have been omitted here.

Earth System modelling, economics, and classical mechanics. Most of the different regions can readily be found in most models for either most or at least selected parameter settings. A notable exception are the "eddies" which, due to their circular nature, can be expected to occur much more rarely in real-world, non-conservative systems, especially when thermodynamic or otherwise irreversible processes are involved, such as soil degradation. Example 3.4, however, illustrates how eddies may occur in coevolutionary systems and might incentivize management cycles that lead to undampened periodic ups and downs. It must remain an open question here whether this effect might be an additional explanation for empirically observable cycles such as business or resource cycles when management is involved.

The introduced concepts have then been used to point out a number of qualitatively different decision problems, the glade, lake, port, harbour, and dock dilemmas. In our opinion, one particularly nasty form of decision problem is the lake dilemma, where one has to choose between uninterrupted desirability and eventual safety, and Example 3.2 indicates that this dilemma may easily occur at least in ecological systems or other multistable systems with a sunny attractor and another one slightly in the dark. Since the transformation of socio-metabolic processes or complex industrial production systems may resemble the soil transformation of Example 3.2, one may also expect the lake dilemma to occur in the socio-metabolic and economic subsystems of the Earth, e.g., in the context of a great transformation leading to decarbonisation of the world's energy system. The form of lake seen near the saddle point in the pendulum Example 3.5 can also occur in other nonlinear oscillators, e.g. the Duffing oscillator or models of glacial cycles that resemble it such as Saltzman et al. (1982); Nicolis (1987), when a management option exists that has a slightly shifted saddle point. This indicates that the lake dilemma may also occur in purely physical subsystems of the Earth System.

We argue that our concepts may be especially useful in the context of the current debate about Planetary Boundaries (PBs), a possible Safe and Just Operating Space (SAJOS) for humanity, and the necessary socio-economic transitions to reach it or stay in it. We suggest that the region delimited by some identified set of Planetary Boundaries in the sense of Rockström et al. (2009a) and Steffen et al. (2015) and some similar socio-economic limits, e.g., those relating to the United Nations sustainable development goals (Raworth, 2012), should be interpreted in our framework as a natural choice for the desirable region X^+, although their definitions already contain some reasoning about the consequences for the respective subsystems when the boundaries are violated. Such boundaries might be called the *Ultimate Planetary Boundaries (UPBs),* and they are typically defined by some simple thresholds for relevant indicators as in Rockström et al. (2009a); Steffen et al. (2015), not taking into account the *overall* system's inherent dynamics much. In this sense, UPBs are typically "non-interacting". Based on the UPBs, one may then try to identify one or more smaller shelter regions S that can be considered a Safe and Just Operating Space (SAJOS) in the sense that, once there, no further large-scale management in the form of global policies is necessary to stay within the limits for all times (or at least for a sufficiently long planning horizon). The borders of these shelters are also a form of PBs but are much more restrictive than the UPBs we started with, and we suggest to call them *Safe Planetary Boundaries (SPBs).*

If it turns out that the current state of the Earth is outside the shelters, one should then aim next at trying to decide whether it is in the upstream. If so, knowledge about whether it is in a glade or lake or not, and which safe docks can be stably reached will be necessary in order to choose a management path. In the glade case, one can still reach the shelter without ever violating the UPBs by appropriate management, hence we suggest to call the border of shelters and glades together the *Provident Planetary Boundaries (PPBs).*

In the lake case, one has to decide instead whether a temporary violation of the UPBs can be justified by the eventual safety of the shelters. In addition, a port dilemma may necessitate a decision between higher desirability and higher flexibility at this point. Only after these qualitative decisions it seems advisable to optimize the chosen type of management pathway by means of more traditional control and optimization theory, hopefully using accurate enough quantitative estimates of the involved options, costs, and benefits. Once in the shelters, one may start caring about improving the state further by moving between docks to either improve desirability or flexibility, but this may require a risky temporary passage through a sunny but unsafe region (which poses a dock dilemma) or even a passage trough the dark (which poses a harbour dilemma). Of course, many combinations of these qualitative and quantitative criteria may appear in the actual global decision process, e.g., in the form of lexicographic preferences, decision trees, or more sophisticated welfare measures or other quantitative objective functions that take the topology suitably into account[1].

If we are not in the "upstream" of the Earth System, prospects are worse. Violating the limits can then only be avoided by management, either eventually forever (if in the downstream), or only repeatedly but with repeated violations occurring (if in the eddies), or even only for a limited time with an ultimate descent into the undesired region (if in the abysses or already in the trench). We suggest to call the upstream borders the *No Regrets Planetary Boundaries (NRPBs).*

If the diagnosis reads "eddy", "abyss" or "trench", one may repeat the analysis with a less ambitious, "second best" definition of the desirable region by choosing less restrictive UPBs, or revert to quantitative optimization, e.g., to mini-

[1] and that may relate to some form of market (or other game theoretic) equilibrium or else be governed by some suitable policy intruments, as kindly suggested by an anonymous referee.

mize some damage function along the system's trajectory. On the other hand, as long as one is in the "manageable region" M (shelters, glades, lakes, and backwaters), the UPBs need never be transgressed if managed wisely, hence we propose to call the borders of M the *Foresighted Planetary Boundaries (FPBs).*

This completes our suggested hierarchy of PBs from the relatively looser UPBs via the successively narrower FPBs and NRPBs, then the PPBs, to the narrowest SPBs that define the SAJOS. While UPBs are "non-interacting", FPBs, PPBs, NRPBs, and SPBS will typically have a more complex geometry in the system's state space and are thus "interacting boundaries". This means that they cannot be expressed as simple "threshold" for individual indicators but as conditional thresholds for several indicators that depend on each other as shown by the curved region boundaries in the examples, e.g., in the carbon cycle model of (Anderies et al., 2013) in Sec. 3.1. Obviously, the real world is less black and white than suggested by the idealised division into "desirable" and "undesirable", so the actual location of these bounds will in reality be somewhat vague, but this does not change the fact that the different bounds and regions represent qualitatively different states of the system, not just quantitative shades of grey.

It should be noted that one strategy to decide the dilemmas described throughout this work is to follow certain "sustainability paradigms" such as those suggested by Schellnhuber (1998). For example, the "pessimization paradigm" is based on the basic precautionary principle of "avoiding the worst" and, hence, can be interpreted as suggesting to stay in or aim for the shelter. In this way, the "pessimization paradigm" decides the glade and lake dilemmas in favour of safety. In turn, the "optimization paradigm" could be interpreted to decide all but the harbor dilemma in favor of uninterrupted or (eventually) higher desirability. The "stabilization paradigm", which seems to fit best the popular notions of "Sustainable Development", reflecting a "longing for stable equilibria" in the coevolutionary dynamics of human societies and the biophysical Earth System (Schellnhuber, 1998), might imply staying in a lake favouring uninterrupted desirability over eventual safety in the sense of this work. Finally, the "equitization paradigm" might imply choosing higher flexibility, e.g., in terms of a larger set of remaining options for future generations in the sense of intergenerational justice, in all dilemmas but the lake dilemma. As also argued by Schellnhuber (1998), the remaining "standardization paradigm" is entirely based on static choices of norms or development corridors instead of dynamical systems or "geocybernetic" principles and, hence, cannot directly decide any of the dilemmas. However, this paradigm can be viewed as a way for identifying desirable domains in the Earth System's state space in the first place and, thereby, facilitate a subsequent topological classification of state space structure.

Contemplating sustainability paradigms gives rise to other relevant qualitative decision problems. For what might be called an "optimization/pessimization dilemma", consider the debate on geoengineering by solar radiation management (Lenton and Vaughan, 2009; Vaughan and Lenton, 2011) as a strategy for averting some of the consequences of global climate change that are induced by anthropogenic emissions of greenhouse gases (Stocker et al., 2013). According to the recent update of the planetary boundary framework by Steffen et al. (2015) and the corresponding definition of desirability (see Sec. 1.1), the Earth System is currently in the dark region of its state space, because core planetary boundaries such as those related to climate change and biosphere integrity have likely already been transgressed. Following current assumptions on the feasibility of management options (Edenhofer et al., 2014), assume further that the Earth System is currently in the dark upstream. In this situation, efforts for mitigation of greenhouse gas emissions, e.g., by means of global energy market regulations, as well as conservation and restoration of biosphere integrity would correspond to navigating the Earth System from the dark upstream towards the shelters following the "pessimization paradigm". In turn, massive investments in solar radiation management as an alternative to mitigation could be seen as manoeuvring the Earth System into the glades or lakes going along with a severe loss of resilience, since interruption of these efforts due to global crisis or technological failure would lead to very rapid and catastrophic climate change (Barrett et al., 2014). In short, starting in the dark upstream, does one choose to navigate to a glade or lake because this appears economically cheaper on the shorter term or politically more feasible ("optimization paradigm") or does one aim for the shelters rightaway, even if this is more expensive on the shorter term ("pessimization paradigm")? Note, however, that geoengineered Earth System states within the glades or lakes would be expected to have a considerably reduced desirably in the long-term compared to the shelters, since current proposals for solar radiation management can only control a very small set of Earth System properties such as global mean temperature, while regional temperature patterns and the hydrological cycle would change strongly (Kleidon and Renner, 2013; Kleidon et al., 2015), going along with corresponding climate impacts.

We hope that the theoretical considerations outlined here may be of some help to sharpen the important debate of how a transition to a safe desirable state of the Earth System can be managed. To this end, future studies should apply the proposed framework for comparing different Earth System governance strategies in the form of various management options (e.g., mitigation of greenhouse gas emissions vs. geoengineering) and different notions of desirability (e.g., resemblance of a Holocene-like state or satisfaction of certain standard of human well-being) in terms of their feasibility and resilience. Furthermore, the structural stability of future development pathways generated by Integrated Assessments Models through optimizing utility functions based on certain notions of human well-being could be evaluated. For achieving these aims, performant computer algorithms need to be developed for automatically generating the proposed topological charts also for higher-dimensional Earth System models given a set of management options and desirability criteria, e.g., building on algorithms from viability theory (Frankowska and Quincampoix, 1990), the graph-theoretical analysis of phase space transition networks (Padberg et al., 2009), and flow networks from fluid dynamics (Ser-Giacomi et al., 2015; Froyland and Padberg-Gehle, 2015). While the examples discussed in this work have been limited to two dynamical variables for facilitating the visualization of the corresponding topological charts, investigation of more detailed models of Earth System dynamics calls for advanced visualization techniques (Nocke et al., 2015) as well as the application and further development of quantitative measures of the size (Menck et al., 2013; Hellmann et al., 2015; van Kan et al., 2015) and shape (Mitra et al., 2015) of the phase space regions of interest. The fact that the introduced state space partitions depend on qualitative rather than quantitative properties of states may also make them a natural tool for the analysis of complex but qualitative or "generalized" models in the spirit of Kuipers (1994); Petschel-Held et al. (1999) or Lade et al. (2013, 2015b, a).

Appendix A: Formal derivation of partitions and properties

We use sloppy set theoretic notation when no confusion arises: union $A+B=A\cup B$, difference $A-B=A\setminus B$, power set $2^A=\{B\subseteq A\}$. Proofs only require an understanding of general topological spaces, in particular of openness and continuity, but not of any higher-level concepts from differential topology or the like.

A1 Assumptions and notation

For a more formal treatment than in the main text, we assume a *manageable dynamical system with desirable states,* made of the following ingredients:

A *state space* $X\neq 0$ with some Hausdorff topology $\mathcal{T}\subseteq 2^X$ (i.e., a system of open sets that separate each two points) on it whose elements we call *states* or *points* (e.g., $X\subseteq\mathbb{R}^n$ with Euclidean topology). X may be compact or unbounded, finite- or infinite-dimensional, etc.

A flow (= deterministic continuous-time autonomous dynamical system) on X (e.g., a model of human-nature coevolution or any other Earth System model) given by a family of continuous ("business-as-usual" or) *default trajectories* $\tau_x:[0,\infty)\to X$ with $\tau_x(0)=x$ and $\tau_{\tau_x(t)}(t')=\tau_x(t+t')$ for all initial conditions $x\in X$ and all relative time points $t,t'\geqslant 0$. We don't require further smoothness properties of the flow, like differentiability, to avoid having to assume a richer topological structure for X than just a general topological space, and to avoid unnecessarily complicated notions and familiarity with, e.g., differential geometry. Although flows are often represented by ordinary differential equations, their solutions are sometimes not unique, hence our notion of flow is in terms of trajectories instead, to allow us to distinguish, e.g., a 1D flow with $\dot{x}=\sqrt{x}$ and $\tau_0(t)\equiv 0$ from the flow that has also $\dot{x}=\sqrt{x}$ but $\tau_0(t)=t^2/4$.

An open nonempty set $X^+\in\mathcal{T}$ of desirable states, called the *sunny region,* e.g., defined by means of some notion of "tolerable E&D window" (Schellnhuber, 1998). We call the complement $X^-=X-X^+\neq 0$ the *dark (region).* We require openness for convenience so that infinitesimal perturbations can't lead from sunny to dark part, and trajectories cannot touch the sunny region without entering it for a strictly positive amount of time. Although in most of our examples, X^+ is a simply shaped, connected, convex, and often compact set, none of these properties is required for the theory presented here except topological openness.

To represent "management options", a family of nonempty sets $\mathcal{M}_x$ of *admissible trajectories* from each $x\in X$ that includes τ_x and is closed under switching between trajectories at any time, i.e., if $\mu\in\mathcal{M}_x$, $t>0$, $x'=\mu(t)$, and $\mu'\in\mathcal{M}_{x'}$, then the trajectory defined by $\mu''(t'')=\mu(t)$ for $t''\leqslant t$ and $\mu''(t'')=\mu'(t''-t)$ for $t''>t$ is also in $\mathcal{M}_x$. This requirement corresponds to the so-called semigroup axiom of mathematical control theory (Sontag, 1998). Note that we do not allow any explicit time dependency of flow or management, but such dependencies can as usual be encoded by including time as a state variable. Also, if management can change a parameter of the model, that parameter has to be transformed to a (slow) state variable with zero default dynamics of its own to meet our framework.

A2 Open invariance, sustainability, and stable reachability

The *invariant open kernel* of a set $A\subseteq X$, denoted $A^{\iota o}$, is the largest open subset of A that contains the default trajectories of all its own points. Its existence and uniqueness is nontrivial and will be proved below. Note that $A^{\iota o}$ may be empty. Each (topologically) connected component of $S=(X^+)^{\iota o}$ is called an individual *shelter.*

We call an open set $A\in\mathcal{T}$ *sustainable* iff for all $x\in A$, there is $\mu\in\mathcal{M}_x$ with $\mu(t)\in A$ for all $t\geqslant 0$. Again, the openness requirement ensures a minimal form of stability against small perturbations. The *sustainable kernel* of a set $A\subseteq X$, denoted $A^{\mathcal{S}}$, is the largest sustainable open subset of A Again, existence and uniqueness will be proved below. In Viability Theory (Aubin, 2001) $A^{\mathcal{S}}$ roughly corresponds to the "viability kernel" of A, see the discussion in Supplement 3. Also $A^{\mathcal{S}}$ may be empty.

Lemma 1 (Existence and uniqueness)**.** *For all $A \subseteq X$:*

1. *There is a unique largest (default-trajectory-) invariant and open subset $A^{\iota\circ} \subseteq A$, containing all other such sets.*
2. *Every invariant and open set is sustainable. In particular, S is.*
3. *There is a unique largest sustainable subset $A^{\mathcal{S}} \subseteq A$ with $A^{\mathcal{S}} \supseteq A^{\iota\circ}$, containing all other such sets.*

Proof.

1. Let $\mathcal{I}(A)$ be the system of all open subsets $B \subseteq A$ for which $\tau_x(t) \in B$ for all $x \in B, t > 0$. The proposition is proved by showing that $\mathcal{I}(A)$ is a *kernel system*, i.e., contains the empty set (which is trivial) and contains the union $\bigcup \mathcal{B}$ of any of its subsets $\mathcal{B} \subseteq \mathcal{I}(A)$. The latter follows from the fact that the system of all open sets, $\mathcal{T}$, is a kernel system by definition, and if $x \in \bigcup \mathcal{B}$, then $x \in B \in \mathcal{B}$, hence $\tau_x(t) \in B \subseteq \bigcup \mathcal{B}$ for all $t > 0$. Now $A^{\iota\circ} = \bigcup \mathcal{I}(A) \in \mathcal{I}(A)$.
2. Since $\tau_x \in \mathcal{M}_x$.
3. Similarly, the system $\mathcal{S}(A)$ of all sustainable subsets $B \subseteq A$ is a kernel system: If $x \in \bigcup \mathcal{B}$, then $x \in B \in \mathcal{B}$, hence there is $\mu \in \mathcal{M}_x$ with $\mu(t) \in B \subseteq \bigcup \mathcal{B}$ for all $t > 0$. Now $A^{\mathcal{S}} = \bigcup \mathcal{S}(A) \in \mathcal{S}(A)$. 2. implies $A^{\mathcal{S}} \supseteq A^{\iota\circ}$.

Q.E.D.

Next, we introduce a suitable notion of stable reachability to overcome two problems with the classical notion of (plain) reachability known from control theory, where a state y is reachable from another state x iff it lies on some admissible trajectory starting at x (Sontag, 1998).

First, we want a stable fixed point y of the default dynamics to be counted as stably reachable from a (sufficiently small) neighbourhood of itself although one might only get arbitrarily close to y instead of getting to y in finite time. Second, we want stable reachability to imply that small perturbations along the way can't render the target unreachable. To solve this conceptual task in a mathematically convenient way, we define stable reachability here via the following binary relation between sets. We call an open set $C \in \mathcal{T}$ a *forecourt* for some set $Y \subseteq X$, denoted $C \rightsquigarrow Y$, iff one can approach Y arbitrarily closely from everywhere in C without leaving C, or, more precisely, iff for all $x \in C$, there is $\mu \in \mathcal{M}_x$ so that for all open sets $Z \in \mathcal{T}$ with $Z \supseteq Y$, there is $t > 0$ with $\mu(t) \in Z$ and $\mu(t') \in C$ for all $t' \in [0,t]$. Now, for a state $x \in X$ and some set $A \subseteq X$, we say that another state $y \in X$ or another set $Y \subseteq X$ are *stably reachable from x through A*, denoted $x \rightsquigarrow_A y$ or $x \rightsquigarrow_A Y$, iff x is in some subset of A that is a forecourt for $\{y\}$ or Y, respectively. The set of states from where Y can be stably reached through A is denoted $(\rightsquigarrow_A Y)$. (This is a stable version of what Aubin (2001) would call a "capture basin" of Y). Note that in these definitions, the order in which the logical quantifiers "for all" and "there exists" appear is critical for some of the resulting properties. If Y is open, the definitions can be somewhat simplified:

Proposition 1 (Stable reachability)**.**
For all $A, A', C, Y, Z \subseteq X$ and $x, y, z \in X$:

1. *If Y is open, then (i) $C \rightsquigarrow Y$ iff for all $x \in C$, there is $\mu \in \mathcal{M}_x$ so that there is $t > 0$ with $\mu(t) \in Y$ and $\mu(t') \in C$ for all $t' \in [0,t]$; and (ii) $x \rightsquigarrow_A Y$ iff there is and open $C \subseteq A$ with $x \in C$ and for all $x' \in C$, there is $\mu \in \mathcal{M}_{x'}$ so that there is $t > 0$ with $\mu(t) \in Y$ and $\mu(t') \in C$ for all $t' \in [0,t]$.*
2. *If $x \rightsquigarrow_A Y$, then x is in the interior (= largest open subset) of A, A°, and there is an open set $B \ni x$ with $x' \rightsquigarrow_A Y$ for all $x' \in B$. Hence, each set of the form $(\rightsquigarrow_A Y)$ is open.*
3. *Transitivity:*

   $$x \rightsquigarrow_A y \rightsquigarrow_{A'} Z \Longrightarrow x \rightsquigarrow_{A+A'} Z,$$
   $$x \rightsquigarrow_A y \rightsquigarrow_{A'} z \Longrightarrow x \rightsquigarrow_{A+A'} z.$$

   In particular, $\rightsquigarrow_A$ is a transitive (but not necessarily reflexive) relation.
4. *If A is open, it is stably reachable from each of its elements. In particular, since $S = (X^+)^{\iota\circ} \subseteq (X^+)^{\mathcal{S}} = M$ is open, S is also included in $U = (\rightsquigarrow_X S)$.*

Proof.

1. (i) Assume $C \rightsquigarrow Y \in \mathcal{T}$ and let $x \in C$. Then, by definition of forecourts, there is $\mu \in \mathcal{M}_x$ so that for all open sets $Z \in \mathcal{T}$ with $Z \supseteq Y$, there is $t > 0$ with $\mu(t) \in Z$ and $\mu(t') \in C$ for all $t' \in [0,t]$. Since Y is open, it is such a Z, proving the first direction.

   For the other direction, assume that for all $x \in C$, there is $\mu \in \mathcal{M}_x$ so that there is $t > 0$ with $\mu(t) \in Y$ and $\mu(t') \in C$ for all $t' \in [0,t]$. Let $x \in C$, choose such a $\mu \in \mathcal{M}_x$ and $t > 0$, and let $Z \in \mathcal{T}$ with $Z \supseteq Y$ be an open set. Then $\mu(t) \in Y \subseteq Z$ as required.

   (ii) By definition of stable reachability, $x \rightsquigarrow_A Y$ iff there is an open $B \subseteq A$ with $x \in B \rightsquigarrow Y$. By (i), $B \rightsquigarrow Y$ iff for all $x' \in B$, there is $\mu \in \mathcal{M}_{x'}$ so that there is $t > 0$ with $\mu(t) \in Y$ and $\mu(t') \in B$ for all $t' \in [0,t]$.
2. Assume $x \rightsquigarrow_A Y$. Then $x \in X$ for some open $B \subseteq A$, hence $x \in B \subseteq A^\circ$. Also, $B \rightsquigarrow Y$ hence $x' \rightsquigarrow_A Y$ for all $x' \in B$. Hence $(\rightsquigarrow_A Y)$ contains an open neighbourhood of each of its points and is thus open itself.
3. We show this by concatenating suitably chosen admissible trajectories between points close to x, y, Z. Let $x \rightsquigarrow_A y \rightsquigarrow_{A'} Z$, choose open sets $B \subseteq A$, $B' \subseteq A'$ with

$x \in B \rightsquigarrow \{y\}$ and $y \in B' \rightsquigarrow Z$, and put $B'' = B + B' \subseteq A + A'$, then $x \in B''$ and B'' is open. To show that $B'' \rightsquigarrow Z$, we let $x'' \in B''$ and show that there is $\mu \in \mathcal{M}_{x''}$ so that for all open $W'' \supseteq Z$, there is $t > 0$ with $\mu(t) \in W''$ and $\mu(t') \in B''$ for all $t' \in [0,t]$.

If $x'' \in B'$, there is such a μ with $\mu(t') \in B' \subseteq B''$ for all $t' \in [0,t]$ since $B' \rightsquigarrow Z$.

If $x'' \notin B'$ instead, $x'' \in B \rightsquigarrow \{y\}$, hence we find $\nu \in \mathcal{M}_{x''}$ so that for all open $W \supseteq \{y\}$, there is $t > 0$ with $\nu(t) \in W$ and $\nu(t') \in B$ for all $t' \in [0,t]$. Since B' is such a W, we find $t'' > 0$ with $\nu(t'') \in B'$ and $\nu(t') \in B$ for all $t' \in [0,t'']$. For $y' = \nu(t'') \in B' \rightsquigarrow Z$, we then find $\nu' \in \mathcal{M}_{x''}$ so that for all open $W'' \supseteq Z$, there is $t > 0$ with $\nu'(t) \in W''$ and $\nu'(t') \in B'$ for all $t' \in [0,t]$. Now define μ by putting $\mu(t') = \nu(t')$ for $t' \in [0,t'']$ and $\mu(t') = \nu'(t' - t'')$ for $t' \geqslant t''$. Then $\mu \in \mathcal{M}_{x''}$ because of our assumptions on $\mathcal{M}$, and for all open $W'' \supseteq Z$, there is $t > 0$ with $\nu'(t) \in W''$ and $\nu'(t') \in B + B' = B''$ for all $t' \in [0,t]$, as required.

The z case follows from putting $Z = \{z\}$. Transitivity is the special case of $A' = A$.

4. For $x \in A \in \mathcal{T}$, we show $x \rightsquigarrow_A A$ by showing $A \rightsquigarrow A$. Let $x' \in A$. By 1., we have to find $\mu \in \mathcal{M}_{x'}$ and $t > 0$ with $\mu(t') \in A$ for all $t' \in [0,t]$. Since A is open and $\tau_{x'}$ is continuous, $\tau_{x'}$ is such a μ.

Q.E.D.

A3 Partitions

A topologically connected component of $\Theta = X - (\rightsquigarrow_X X^+)$, $\Upsilon = \overline{\{x \in X \,|\, \forall \mu \in \mathcal{M}_x \,\exists t \geqslant 0 : \mu(t) \in \Theta\}} - \Theta$, or $E = X - U - D - \Upsilon - \Theta$ will be called an individual *trench, abyss,* or *eddy,* and the latter two typically have sunny and dark parts. Some further properties of these introduced partition sets are:

Proposition 2 (Main cascade)**.**

1. *$U = (\rightsquigarrow_X S)$ and the union $D + U = (\rightsquigarrow_X M)$ are open, $\Theta = X - (\rightsquigarrow_X X^+)$ and $\Upsilon + \Theta$ are closed, the union $E + D + U = X - \Upsilon - \Theta$ is open, and the system $\{U, D, E, \Upsilon, \Theta\}$ forms a partition of X.*
2. *For all $u \in U, d \in D, e \in E, y \in \Upsilon, \theta \in \Theta$, we have $u \not\rightsquigarrow_X d \not\rightsquigarrow_X e \not\rightsquigarrow_X y \not\rightsquigarrow_X \theta$.*
3. *If $W = \emptyset$, also $D = \emptyset$.*

Proof.

1. Openness follows from Prop. 1,1., the partition covers X by definition of E, and the only nontrivial disjointness is that between the open set $D + U = (\rightsquigarrow_X M)$ and the closed set $\Upsilon + \Theta = \overline{\{x \in X \,|\, \forall \mu \in \mathcal{M}_x \,\exists t \geqslant 0 : \mu(t) \in \Theta\}}$. If x is in both sets, there is also $x' \in (\rightsquigarrow_X M) \cap \{x \in X \,|\, \forall \mu \in \mathcal{M}_x \,\exists t \geqslant 0 : \mu(t) \in \Theta\}$, but then there is $\mu'_x \in \mathcal{M}_x$, $t' > 0$ with $\mu'_x(t') \in M$, and by definition of M there is then also some $\mu \in \mathcal{M}_x$ with $\mu(t) \in X^+$ for all $t \geqslant t'$. But, by assumption, there is $t \geqslant 0$ with $\mu(t) \in \Theta$. Since $\Theta \cap X^+ = 0$, we have $t < t'$, but by definition of Θ, this contradicts $\mu(t') \in X^+$. Hence such an x cannot exist.
2. Because of transitivity and 1., $d \rightsquigarrow_X u \in U = (\rightsquigarrow_X S)$ would imply $d \rightsquigarrow_X S$ and thus $d \in U \cap D = \emptyset$; $e \rightsquigarrow_X d \in D = (\rightsquigarrow_X M) - U$ would imply $e \rightsquigarrow_X M$ and thus $e \in (U + D) \cap E = \emptyset$. If one could reach the eddies from the abysses, one could avoid the trenches: Assume $y \rightsquigarrow_X e \notin \Upsilon + \Theta = \overline{\{x \in X \,|\, \forall \mu \in \mathcal{M}_x \,\exists t \geqslant 0 : \mu(t) \in \Theta\}}$. Since the latter is closed, its complement is open, so there is $\mu \in \mathcal{M}_y$ and $t > 0$ with $\mu(t) \notin \Upsilon + \Theta$. For $x = \mu(t)$, we find $\mu' \in \mathcal{M}_x$ and $t'' > 0$ with $\mu'(t') \notin \Theta$ for all $t' > t''$. Concatenating μ with μ' gives a similar member of $\mathcal{M}_y$, in contradiction to $y \in \Theta$. Finally, if $\theta \rightsquigarrow_X y$ and $\theta \in \Theta$, then $y \in \Theta$ by definition of Θ, hence $y \notin \Upsilon$.
3. This follows from $(\rightsquigarrow_X M) - U = D = (\rightsquigarrow_X W)$.

Q.E.D.

Note that in the (pathological) *no-management case* in which $\mathcal{M}_x = \{\tau_x\}$, the upstream $U = (\rightsquigarrow_X S)$ is basically (i.e., up to boundary effects due to our openness requirement) the basin of attraction of S, the downstream $D = (\rightsquigarrow_X M) - (\rightsquigarrow_X S)$ is then empty, the trenches basically equal the invariant kernel of X^-, the abysses basically equal the rest of the basin of attraction of the trenches, and the eddies is basically the union of those trajectories that will forever alternate between X^+ and X^-. In that case also some of the finer regions may coincide or be empty, and one can represent their relationship also by means of *symbolic dynamics* (Beim Graben and Kurths, 2003): Assign each state x a symbolic sequence representing the sequence of its trajectory's transitions between the sunny $(+)$ and dark $(-)$ regions, and use the wildcard $*$ to denote repetitions of zero or more symbols. Then (up to peculiarities that may occur for boundary states) $S = M = (+)$, $U^- = (-)(+-)^*(+)$, $U^{(+)} = (+-)(+-)^*(+)$, $G = L = D = \emptyset$, $E^+ = (+-)^\infty$, $E^- = (-+)^\infty$. $\Upsilon^+ = (+)(-+)^*(-)$, $\Upsilon^- = (-+)(-+)^*(-)$, and $\Theta = (-)$.

To formally define the ports-and-rapids partition, we say that a set $P \subseteq X$ is *portish* iff it has $x \rightsquigarrow_X y$ for all $x, y \in P$, is topologically connected, and does not intersect two different eddies, abysses, or trenches. A maximal portish set is called a *port.*

We show below that each two ports are disjoint, each port is completely contained in one of the sets $U, D, E, \Upsilon^-, \Theta$, none can intersect Υ^+, each *returnable* state (i.e., an x with

$x \rightsquigarrow_X x$) is in a port, but no *transitional* state (x with $x \not\rightsquigarrow_X x$) is.

In the pendulum example of Fig. 8, the returnable points are those in $U+D$ because of the periodic frictionless default flow and the possibility of counteracting small perturbations by braking or accellerating at some later point of the perturbed trajectory. In the eddies and below, this is not possible after an accelerating perturbation, hence those regions are transitional. In the plant types example of Fig. 5, there are also transitional regions, e.g. to the top and right where all admissible trajectories lead down and left; and in the technological change example of Fig. 6 all points are transitional because of the positive growth of the knowledge stocks.

To extend the system $\mathcal{P}$ of all ports into a partition of all of X that is finer than the main cascade $\mathcal{C}$, we say that two non-port states x,y are *port-equivalent* iff they are in the same member of $\mathcal{C}$, do not lie in two different eddies, abysses, or trenches, and if $x \rightsquigarrow_X P \Leftrightarrow y \rightsquigarrow_X P$ and $P \rightsquigarrow_X x \Leftrightarrow P \rightsquigarrow_X y$ for all $P \in \mathcal{P}$. Each maximal topologically connected set of port-equivalent states is now called a *rapid*. This ensures that not only U and D are partitioned into ports and rapids, but so is each individual eddy, abyss, and trench. The ports and rapids together form the *ports and rapids partition*, $\mathcal{PR}$, which is finer than $\mathcal{C}$.

A set $H \subseteq X$ is *harbourish* iff it has has $x \rightsquigarrow_{X^+} y$ for all $x,y \in H$, is topologically connected, does not intersect two different lakes, eddies, or abysses, and does not intersect two different connected components of $S+G$. A maximal harbourish set is called a *harbour.* Let $\mathcal{H}$ be the system of all harbours. Two non-harbour states $x,y \in X^+$ are *harbour-equivalent* iff they are in the same member of $\{S+G, L, U^{(+)}, W, D^{(+)}, E^+, \Upsilon^+\}$, do not lie in two different lakes, eddies, or abysses, do not lie in two different connected components of $S+G$, and if $x \rightsquigarrow_{X^+} H \Leftrightarrow y \rightsquigarrow_{X^+} H$ and $H \rightsquigarrow_{X^+} x \Leftrightarrow H \rightsquigarrow_{X^+} y$ for all $H \in \mathcal{H}$. Each maximal topologically connected set of harbour-equivalent states is called a *channel* and lies completely in either one port or one rapid (see below for a proof), hence The resulting *harbours and channels partition* of X^+, $\mathcal{HC}$, is finer than $\mathcal{PR}$.

A set $O \subseteq X$ is *dockish* iff it has $x \rightsquigarrow_S y$ for all $x,y \in O$, is topologically connected and does not intersect two different shelters. A maximal dockish set is called a *dock.* Let $\mathcal{O}$ be the system of all docks. Two non-dock states $x,y \in S$ are called *dock-equivalent* iff they belong to the same shelter and $x \rightsquigarrow_S O \Leftrightarrow y \rightsquigarrow_S O$ and $O \rightsquigarrow_S x \Leftrightarrow O \rightsquigarrow_S y$ for all $O \in \mathcal{O}$. Each maximal topologically connected set of dock-equivalent states is called a *fairway* and lies completely in either one harbour or one channel, hence the resulting *docks and fairways partition* of S, $\mathcal{OF}$, is finer than $\mathcal{HC}$.

Proposition 3 (Ports, rapids, harbours, etc.)**.**

1. *Each two ports [or harbours or docks] are disjoint.*
2. *Each port lies completely in one of $U, D, E, \Upsilon^-, \Theta$, no port intersects Υ^+.*
3. *Each harbour [or dock] lies completely in one port [or harbour].*
4. *Each channel [or fairway] lies completely in one member of $\mathcal{PR}$ [or $\mathcal{HC}$].*
5. *These partitions are successive refinements of each other: $\mathcal{C}, \mathcal{PR}, \mathcal{HC}, \mathcal{OF}$.*
6. *If a harbour H intersects some of the regions $S+G$, L, U^+, W, or D^+, it is already completely contained in that region.*

Proof.

1. Assume $y \in A \cap A'$ for two different maximal portish [or harbourish or dockish] sets A, A' and put $B = A + A'$. But then B is itself portish [or harbourish or dockish] because stable reachability is transitive. This contradicts the maximality of A and A'.

2. By Prop. 2,2., if $x \rightsquigarrow_P y \rightsquigarrow_P x$ then x and y they must belong to the same member of $\mathcal{C}$, hence each port lies completely in one of them.

   To show that a port $P \subseteq \Upsilon$ is already in Υ^-, assume $x \in P \cap \Upsilon^+ \subseteq X^+ \in \mathcal{T}$. We will now construct a contradiction by constructing an admissible trajectory from x that avoids Θ forever. Since $x \rightsquigarrow_X x$ and X^+ is open, there is an open set $A \subseteq X^+$ with $y \rightsquigarrow_X x$ for all $y \in A$. Since τ_x is continuous and A open, we find $t_0 > 0$ with $\tau_x(t) \in A$ for all $t \in [0, t_0]$. Let $y = \tau_x(t_0)$ and pick a $\mu \in \mathcal{M}_y$ that returns arbitrarily closely to x. Let $\mathcal{A}$ be the set of all open $A \subseteq X^+$ with $x \in A$, and choose a $t_A > 0$ with $\mu(t_A) \in A$ for all $A \in \mathcal{A}$ (this requires the Axiom of Choice which we will assume here). Let $t_1 = \inf_{A \in \mathcal{A}} \sup_{B \in \mathcal{A}, B \subseteq A} t_B \geqslant 0$. Since $y \in \Upsilon + \Theta$, there is $t' > 0$ with $\mu(t'') \in \Theta$ for all $t'' > t'$, hence $t_A \leqslant t'$ for all $A \in \mathcal{A}$ and thus $t_1 \leqslant t'$. Next we show that $\mu(t_1) = x$. If $\mu(t_1) = y \neq x$, one can choose $A \in \mathcal{A}$ and $C \in \mathcal{T}$ with $y \in C$ and $A \cap C = \emptyset$ (this is the only point where we need the Hausdorff property). Since μ is continuous, there are $t_l < t_1$ and $t_u > t_1$ with $\mu(t') \in C$ for all $t' \in [t_l, t_u]$. By definition of t_1, there is $A' \in \mathcal{A}$ with $\sup_{B \in \mathcal{A}, B \subseteq A'} t_B \in [t_1, t_u]$. Putting $A'' = A \cap A' \in \mathcal{A}$, we then also have $\sup_{B \in \mathcal{A}, B \subseteq A''} t_B \in [t_1, t_u]$, hence there is $B \in \mathcal{A}$ with $B \subseteq A'' \subseteq A$ and $t_B \geqslant t_l$ and hence $\mu(t_B) \in C$ by choice of t_l. But $\mu(t_B) \in B \subseteq A$ by choice of t_B. Hence $\mu(t_B) \in A \cap C = \emptyset$, a contradiction. So $\mu(t_1) = x$ after all. Finally we concatenate $\tau_x[0, t_0]$ and $\mu[0, t_1]$ infinitely many times and get an admissible trajectory from x that avoids Θ forever.

3. Since $\rightsquigarrow_S$ refines $\rightsquigarrow_{X^+}$, which refines $\rightsquigarrow_X$.

4. Since dock-equivalence refines harbour-equivalence, which refines port-equivalence.

5. Follows from 2.–4.

6. This follows directly from the definitions of $S+G,L,U^+,W$, and D^+ by means of $\rightsquigarrow_X$ and $\rightsquigarrow_{X^+}$ and the transitivity of those relations.

Q.E.D.

A4 Remarks

– In general, $A^{\iota\circ}$ may be properly smaller than both the interior $(A^\iota)^\circ$ of the largest invariant subset A^ι of A and the largest invariant subset of A°, $(A^\circ)^\iota$. The three sets can only be shown to be equal under additional smoothness assumptions on τ and $\mu \in \mathcal{M}_x$.

– The set of all states that are stably reachable from x need not be closed or open and need not contain any of the intermediate states that lie on the trajectories $\mu \in \mathcal{M}_x$ used in stable reachability.

– $x \rightsquigarrow_A Y$ does not imply $x \rightsquigarrow y$ for any $y \in Y$, since, after a perturbation, other points in Y may be reachable than before.

– For two points x, y in the same port, harbour, or dock A, one may still not have $x \rightsquigarrow_A y$ since the intermediate states on the trajectories from x to y may not be *stably* reachable from x and thus may not belong to A. In other words, perturbations may still push the system temporarily out of a port, harbour, or dock, but one can then return to the same port, harbour, or dock. For this reason, the directed reachability network is typically acyclic but may contain reachability cycles in pathological situations.

– Any attractor A with the return property (e.g., a stable fixed point or limit cycle, and most strange and chaotic attractors) of the default dynamics lies completely within one port, hence within one member of $\mathcal{C}$. If $A \subseteq X^+$ then already $A \subseteq S$ and A lies completely within one dock.

– The scope of possible connection topologies that may occur as the reachability network of a managed system contains at least all acyclic finite or countably infinite directed graphs, as can be seen by the following construction: given an acyclic directed graph, one can construct a topologically equivalent network of water bowls which are connected by water tubes leading from a dedicated "drain" at the bottom of the source ball to a common entrance at the top of the target ball. Let water flow into all balls without incoming tubes and out of all outgoing tubes through grilles, determining the default dynamics of a small submarine floating in the water. Then assume the submarine can be propelled strongly enough to move freely inside each ball and to each drain, but not strongly enough to leave the ball through the entrance at the top, against the direction of the water flow. By making parts of the balls and tubes opaque and moving some of the drains from the bottom to the sides of the ball, the construction can be extended to show that also all internally consistent three-level acyclic networks can occur as the three-level network of ports, harbours, and docks.

Appendix B: Further examples

B1 One-dimensional potential function

This simple model shows how almost all of the introduced state space regions (except eddies and dark abysses) may already occur in a one-dimensional system $\dot{x} = -df/dx$ that is defined by a potential function $f(x)$ and already for simple desirable regions such as $X^+ =]0, \infty[$, as depicted in Fig. B1.

Our example has a default dynamics along the blue line downwards at a speed proportional to slope, but management is able to move upwards instead on the thin blue lines where the slope is small enough (for $|df/dx| < 3/2$). The chosen undesirable region of $x \leqslant 0$ is indicated in grey. The shelter consists of the two segments just left of point a and it can be stably reached from everywhere properly left of a, hence that whole region constitutes the upstream. The manageable region is the union of shelter, glade, lake, and backwater, and it can be stably reached from everywhere properly left of point b, hence the downstream is the right-open interval from a to b.

That there are no eddies and no dark abysses in this example is typical for systems without any circular flows and with a sufficiently simply shaped X^+.

There are two ports, the two closed intervals where the default flow is slow, one in the upstream and one in the downstream. Note that the latter is only partially contained in the backwater. One rapid lies to the left of the left port, another between the left port and point a, and these two rapids are port-equivalent since both can reach the left but not the right port. Similarly, the right port is surrounded by two port-equivalent rapids. Finally, there is a singleton rapid consisting only of the point a and a last one formed by point b and all that is to the right of it; from these two port-equivalent rapids, no port can be stably (!) reached.

B2 Bifurcations of a directly manageable flow

If a system passes through a bifurcation, the classification of states by the criteria outlined above will typically change. Let us examine some archetypical cases that can occur in

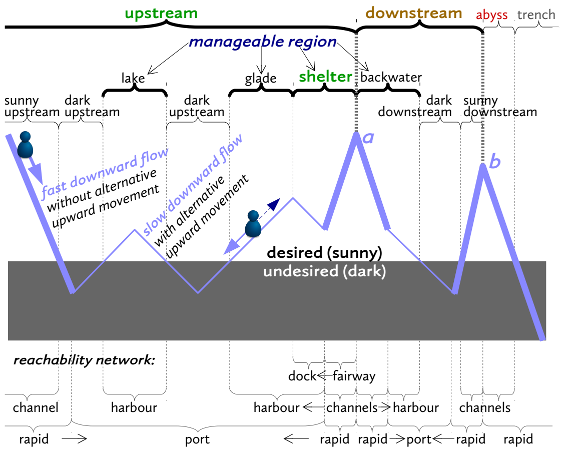

Figure B1. A system moves along the blue line: downward by default (pale blue arrows), but in some regions management can move it in the opposite direction (dark blue arrow) in order to avoid the undesired "dark" region. *Shelters, manageable region, upstream & downstream* (boldface, Section 2.2) and other regions from the *main cascade* (top line, Section 2.3). Regions from the finer *manageable partition* (below, Section 2.4). See Fig. 2 for a systematic summary of these concepts. Bottom: three-level *reachability network* (Section 2.5).

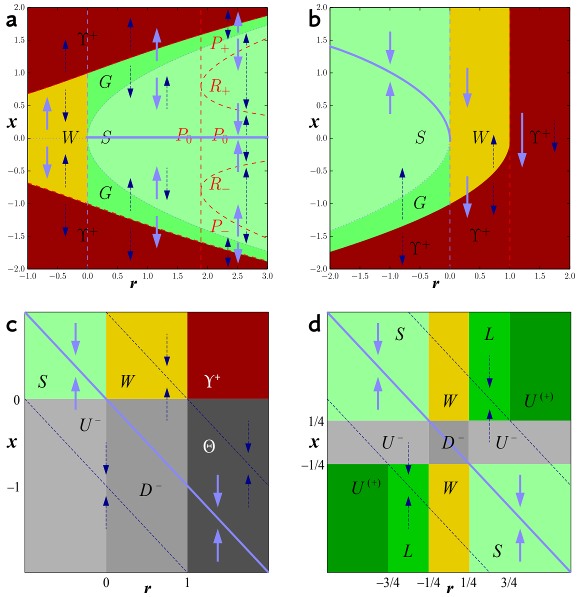

Figure B2. Parameter changes can change the quality of states due to bifurcations. Top-left: backwater/glade bifurcation and later port pitchfork bifurcation caused by a subcritical pitchfork bifurcation of the default flow (similar in the supercritical case). Top-right: glade/backwater/abyss transition caused by a saddle-node bifurcation, with the second critical value marked in red. Bottom-left: shelter/backwater/abyss transition caused by the transition of a stable fixed point into the deep dark. Bottom-right: shelter/backwater/lake/upstream transition caused by the transition of a stable fixed point through a dark strip.

the exemplary case where management can directly affect the flow by changing the default derivative $\dot{x} = F(x)$ of a one-dimensional system by at most one unit, so that the admissible trajectories are those with $\dot{x} \in [F(x)-1, F(x)+1]$. (See Example 3.6 above for the case where management is via changing a parameter instead).

Assume $X^+ = \{|x| < \ell\}$ for some $\ell \gg 1$, and the default flow has a *subcritical pitchfork bifurcation,* say $F(x) = x^3 - rx$, where for $r > 0$ the stable fixed point $x_0 = 0$ is surrounded by two unstable ones at $x_\pm = \pm\sqrt{r}$ and becomes unstable itself for $r \leqslant 0$, as depicted by the solid and dotted pale blue lines in Fig. B2 a). Then for $r > 0$, we have a shelter-and-glade situation with a shelter $S =]-\sqrt{r}, \sqrt{r}[$ and two glades $G =]-g(r), -\sqrt{r}] + [\sqrt{r}, g(r)[$ where $g(r) > \sqrt{r}$ is the upper solution to the equation $F(g(r)) - 1 = 0$, indicating the limit above which also the extreme management with $\dot{x} = F(x) - 1$ cannot move the system downwards (dashed dark blue lines). But for $r \leqslant 0$, the shelter disappears and the glades merge and are converted into a backwater $W =]-g(r), g(r)[$. In both cases, this is surrounded by two sunny abysses $\Upsilon^+ =]-\ell, -g(r)] + [g(r), \ell[$ and two trenches $\Theta =]-\infty, \ell] + [\ell, \infty[$ (outside the depicted area). One may call this transition a *backwater/glade bifurcation.* As an early warning signal of an imminent breakdown of a shelter in such a backwater/glade bifurcation, one may consider the volume of the shelters $\mathrm{Vol}(S)$ in terms of some natural measure on X as a measure of "shelter stability", similar to the concept of basin stability for unmanaged systems without desirable region (Menck et al., 2013; Ji and Kurths, 2014; Schultz et al., 2014; van Kan et al., 2015) and to the recently introduced survivability measure for unmanaged systems with a desirable region (Hellmann et al., 2015).

The port surrounding the unstable fixed point $x = 0$, $P_0 =]-g(r), g(r)[$, where $g(r)$ is the solution to $F(g(r)) + 1 = 0$, eventually also splits in three ports P_0 and $P_\pm$, separated by two rapids $R_\pm$; their borders are depicted by the dashed red lines. But this happens only at a larger value of r, namely at $r = 3/\sqrt[3]{4} \approx 1.9$, after which the two unstable fixed points $x_\pm$ can no longer be reached from each other. The corresponding ports and rapids network has these arrows: $P_- \leftsquigarrow_X R_- \leftsquigarrow_X P_0 \rightsquigarrow_X R_+ \rightsquigarrow_X P_+$. One may call this transition a *port pitchfork bifurcation.*

An interesting case is a *saddle-node bifurcation* such as the one in Fig. B2 b), with $F(x) = -r - x^2$ and a critical parameter value $r = 0$ at which the stable and unstable fixed points at $x = \pm\sqrt{-r}$ collide and disappear. First, at the critical point, the shelter caused by the stable fixed point and its glade are transformed into a backwater. Then, somewhat later (at $r = 1$), the maximally value of $\dot{x}$ achievable by management becomes negative and the backwater ceases to exist so that only the sunny abyss remains. One may call this a *glade/backwater/abyss transition.*

If a stable fixed point approaches and eventually enters deeply into the dark region, this may also be called a form of “bifurcation” that causes a similar transition in the classification of states. If $F(x) = -r - x$ and $X^+ = \{x > 0\}$, as in Fig. B2 c), then again two changes occur: At $r = 0$, the shelter-and-upstream situation of $r < 0$, with $S =]0,\infty[$ and $U^- =]-\infty,0]$, converts into a backwater-and-downstream situation with $W =]0,\infty[$ and $D^- =]-\infty,0]$. Then at $r = 1$, this further converts into an abyss-and-trench situation of $r \geqslant 1$ with $\Upsilon^+ =]0,\infty[$ and $\Theta =]-\infty,0]$. One could thus call this a *shelter/backwater/abyss transition.*

Finally, a transition with three steps is caused if the fixed point passes through a narrower strip of dark, as in Fig. B2 d), where again $F(x) = -r - x$ but now $X^+ = \{|x| > 1/4\}$. Here the shelter is again first transformed into a backwater at $r = -1/4$, but then into a lake L when the fixed point leaves the dark again at $r = +1/4$, and even later into a remaining sunny upstream $U^{(+)}$ once the maximally achievable value of $\dot{x}$ at the upper boundary of the dark, i.e., at $x = 1/4$, becomes negative. We suggest to call this a *shelter/backwater/lake/upstream transition.*

Acknowledgements. This work was conducted in the framework of PIK’s flagship project on Coevolutionary Pathways (COPAN). This paper was developed within the scope of the IRTG 1740 / TRP 2011/50151-0, funded by the DFG / FAPESP. JFD thanks the Stordalen Foundation (via the Planetary Boundary Research Network PB.net) and the Earth League’s EarthDoc program for financial support. We thank Wolfram Barfuss, Boyan Beronov, Nicola Botta, Ingo Fetzer, Vera Heck, Ulrike Kornek, Jürgen Kurths, Steven Lade, Wolfgang Lucht, Finn Müller-Hansen, John Schellnhuber, Ricarda Winkelmann, the COPAN team, and three anonymous referees for fruitful discussions and helpful comments.

References

Anderies, J. M., Carpenter, S. R., Steffen, W., and Rockström, J.: The topology of non-linear global carbon dynamics: from tipping points to planetary boundaries, Environmental Research Letters, 8, 044 048, doi:10.1088/1748-9326/8/4/044048, 2013.

Aubin, J.-P.: Viability Kernels and Capture Basins of Sets Under Differential Inclusions, SIAM Journal on Control and Optimization, 40, 853–881, doi:10.1137/S036301290036968X, 2001.

Aubin, J.-P.: Viability theory, Birkhäuser, 2009.

Aubin, J.-P. and Saint-Pierre, P.: An Introduction to Viability Theory and Management of Renewable Resources, in: Advanced Methods for Decision Making and Risk Management in Sustainability Science, edited by Kropp, J. and Scheffran, J., chap. 2, pp. 43–80, Nova Science Publishers, 2007.

Aubin, J.-P., Bayen, A., and Saint-Pierre, P.: Viability Theory. New Directions, Springer Science & Business Media, 2011.

Ayres, R. U., van den Bergh, J. C., and Gowdy, J. M.: Strong versus weak sustainability: Economics, natural sciences, and ’consilience’, Environmental Ethics, 23, 155–168, 2001.

Barrett, S., Lenton, T. M., Millner, A., Tavoni, A., Carpenter, S., Anderies, J. M., Chapin III, F. S., Crépin, A.-S., Daily, G., Ehrlich, P., et al.: Climate engineering reconsidered, Nature Climate Change, 4, 527–529, doi:10.1038/nclimate2278, 2014.

Beim Graben, P. and Kurths, J.: Detecting subthreshold events in noisy data by symbolic dynamics., Physical Review Letters, 90, 100 602, doi:10.1103/PhysRevLett.90.100602, 2003.

Beven, K. J.: Searching for the Holy Grail of scientific hydrology: Qt = H(S, R, t) A as closure, Hydrology and Earth System Sciences, 10, 609–618, doi:10.5194/hess-10-609-2006, 2006.

Botta, N., Jansson, P., and Ionescu, C.: A computational theory of policy advice and avoidability, http://www.cse.chalmers.se/~patrikj/papers/CompTheoryPolicyAdviceAvoidability_preprint.pdf, 2015.

Brander, J. A. and Taylor, M. S.: The simple economics of Easter Island: A Ricardo-Malthus model of renewable resource use, The American Economic Review, 88, 119–138, 1998.

Bruckner, T. and Zickfeld, K.: Inverse Integrated Assessment of Climate Change: the Guard-Rail Approach, in: International Conference on Policy Modeling (EcoMod2008), 2008.

Carpenter, S. R., Brock, W. A., Folke, C., van Nes, E. H., and Scheffer, M.: Allowing variance may enlarge the safe operating space for exploited ecosystems, Proceedings of the National Academy of Sciences, online first, doi:10.1073/pnas.1511804112, 2015.

Dasgupta, P.: Discounting climate change, Journal of Risk and Uncertainty, 37, 141–169, 2008.

Edenhofer, O., Knopf, B., Barker, T., Baumstark, L., Bellevrat, E., Chateau, B., Criqui, P., Isaac, M., Kitous, A., Kypreos, S., Leimbach, M., Lessmann, K., Magné, B., Scrieciu, v., Turton, H., and Van Vuuren, D. P.: The economics of low stabilization: Model comparison of mitigation strategies and costs, The Energy Journal, 31, 11–48, doi:10.5547/ISSN0195-6574-EJ-Vol31-NoSI-2, 2010.

Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., et al.: Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press Cambridge; New York, NY, 2014.

Folke, C., Carpenter, S. R., Walker, B., Scheffer, M., Chapin, T., and Rockström, J.: Resilience thinking: integrating resilience, adaptability and transformability, Ecology and Society, 15, 20, 2010.

Folke, C., Jansson, Å., Rockström, J., Olsson, P., Carpenter, S. R., Chapin III, F. S., Crépin, A.-S., Daily, G., Danell, K., Ebbesson, J., et al.: Reconnecting to the biosphere, Ambio, 40, 719–738, doi:10.1007/s13280-011-0184-y, 2011.

Frankowska, H. and Quincampoix, M.: Viability kernels of differential inclusions with constraints: algorithms and applications, International Institute for Applied Systems Analysis Working Paper, 1990.

Froyland, G. and Padberg-Gehle, K.: A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data, arXiv preprint arXiv:1505.04583, 2015.

Ganopolski, a. and Rahmstorf, S.: Rapid changes of glacial climate simulated in a coupled climate model., Nature, 409, 153–158, doi:10.1038/35051500, 2001.

Heitzig, J., Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: Node-weighted measures for complex networks with spatially embedded, sampled, or differently sized nodes, The European Physical Journal B, 85, 38, doi:10.1140/epjb/e2011-20678-7, 2012.

Hellmann, F., Schultz, P., Grabow, C., Heitzig, J., and Kurths, J.: Survivability: A Unifiying Concept for the Transient Resilience of Deterministic Dynamical Systems, arXiv preprint arXiv:1506.01257, 2015.

Jaffe, A., Newell, R., and Stavins, R.: Environmental policy and technological change, Environmental and Resource Economics, 22, 41–69, doi:10.1023/A:1015519401088, 2002.

Janssen, R. H. H., Meinders, M. B. J., van NES, E. H., and Scheffer, M.: Microscale vegetation-soil feedback boosts hysteresis in a regional vegetation–climate system, Global Change Biology, 14, 1104–1112, doi:10.1111/j.1365-2486.2008.01540.x, 2008.

Ji, P. and Kurths, J.: Basin stability of the Kuramoto-like model in small networks, The European Physical Journal Special Topics, doi:10.1140/epjst/e2014-02213-0, 2014.

Kalkuhl, M., Edenhofer, O., and Lessmann, K.: Learning or lock-in: Optimal technology policies to support mitigation, Resource and Energy Economics, 34, 1–23, doi:10.1016/j.reseneeco.2011.08.001, 2012.

Keller, K., Hall, M., Kim, S. R., Bradford, D. F., and Oppenheimer, M.: Avoiding dangerous anthropogenic interference with the climate system, Climatic Change, 73, 227–238, doi:10.1007/s10584-005-0426-8, 2005.

Kleidon, A. and Renner, M.: A simple explanation for the sensitivity of the hydrologic cycle to surface temperature and solar radiation and its implications for global climate change, Earth System Dynamics, 4, 455–465, doi:10.5194/esd-4-455-201, 2013.

Kleidon, A., Kravitz, B., and Renner, M.: The hydrological sensitivity to global warming and solar geoengineering derived from thermodynamic constraints, Geophysical Research Letters, 42, 138–144, doi:10.1002/2014GL062589, 2015.

Kreps, D. M.: A Representation Theorem for "Preference for Flexibility", Econometrica, 47, 565–577, doi:10.2307/1910406, 1979.

Kuipers, B.: Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge, MIT Press, Cambridge, MA, 1994.

Lade, S. J., Tavoni, A., Levin, S. A., and Schlüter, M.: Regime shifts in a social-ecological system, Theoretical ecology, 6, 359–372, doi:10.1007/s12080-013-0187-3, 2013.

Lade, S. J., Niiranen, S., Hentati-Sundberg, J., Blenckner, T., Boonstra, W. J., Orach, K., Quaas, M. F., Österblom, H., and Schlüter, M.: An empirical model of the Baltic Sea reveals the importance of social dynamics for ecological regime shifts, Proceedings of the National Academy of Sciences, 112, 11 120–11 125, doi:10.1073/pnas.1504954112, 2015a.

Lade, S. J., Niiranen, S., and Schlüter, M.: Generalized modeling of empirical social-ecological systems, arXiv preprint arXiv:1503.02846, 2015b.

Lenton, T. M. and Vaughan, N. E.: The radiative forcing potential of different climate geoengineering options, Atmospheric Chemistry and Physics, 9, 5539–5561, doi:10.5194/acp-9-5539-2009, 2009.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J.: Tipping elements in the Earth's climate system, Proceedings of the National Academy of Sciences of the United States of America, 105, 1786–1793, doi:10.1073/pnas.0705414105, 2008.

Martin, S.: The cost of restoration as a way of defining resilience: a viability approach applied to a model of lake eutrophication, Ecology And Society, 9, 8, 2004.

Menck, P. J., Heitzig, J., Marwan, N., and Kurths, J.: How basin stability complements the linear-stability paradigm, Nature Physics, 9, 89–92, doi:10.1038/nphys2516, 2013.

Mitra, C., Kurths, J., and Donner, R. V.: An integrative quantifier of multistability in complex systems based on ecological resilience, Scientific Reports, 5, 1–12, doi:10.1038/srep16196, 2015.

Nagy, B., Farmer, J. D., Bui, Q. M., and Trancik, J. E.: Statistical Basis for Predicting Technological Progress, PLoS ONE, 8, 1–7, doi:10.1371/journal.pone.0052669, 2013.

Nicolis, C.: Long-term climatic variability and chaotic dynamics, Tellus A, pp. 1–9, doi:10.3402/tellusa.v39i1.11734, 1987.

Nocke, T., Buschmann, S., Donges, J., Marwan, N., Schulz, H.-J., and Tominski, C.: Review: visual analytics of climate networks, Nonlinear Processes in Geophysics Discussions, 2, 709–780, doi:10.5194/npgd-2-709-2015, 2015.

Padberg, K., Thiere, B., Preis, R., and Dellnitz, M.: Local expansion concepts for detecting transport barriers in dynamical systems, Communications in Nonlinear Science and Numerical Simulation, 14, 4176–4190, doi:10.1016/j.cnsns.2009.03.018, 2009.

Petschel-Held, G., Schellnhuber, H.-J., Bruckner, T., Toth, F. L., and Hasselmann, K.: The tolerable windows approach: theoretical and methodological foundations, Climatic Change, 41, 303–331, doi:10.1023/A:1019080704864, 1999.

Rahmstorf, S., Crucifix, M., Ganopolski, a., Goosse, H., Kamenkovich, I., Knutti, R., Lohmann, G., Marsh, R., Mysak, L., Wang, Z., and a.J. Weaver: Thermohaline circulation hysteresis: a model intercomparison, Geophysical Research Letters, 32, 1–5, doi:10.1029/2005GL023655, 2005.

Raworth, K.: A Safe and Just Space For Humanity: Can we live within the Doughnut?, Oxfam Policy and Practice: Climate Change and Resilience, 8, 1–26, doi:10.5822/978-1-61091-458-1, 2012.

Rockström, J., Steffen, W., Noone, K., and Persson, A.: Planetary boundaries: exploring the safe operating space for humanity, Ecology and Society, 14, 32, 2009a.

Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R. W., Fabry, V. J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., and Foley, J. A.: A safe operating space for humanity., Nature, 461, 472–475, doi:10.1038/461472a, 2009b.

Rougé, C., Mathias, J. D., and Deffuant, G.: Extending the viability theory framework of resilience to uncertain dynamics, and application to lake eutrophication, Ecological Indicators, 29, 420–433, doi:10.1016/j.ecolind.2012.12.032, 2013.

Saltzman, B., Sutera, A., and Hansen, A. R.: A Possible Marine Mechanism for Internally Generated Long-Period Climate Cycles, Journal of the Atmospheric Sciences, 39, 2634–2637, doi:10.1175/1520-0469(1982)039<2634:APMMFI>2.0.CO;2, 1982.

Scheffer, M., Barrett, S., Carpenter, S., Folke, C., Green, A. J., Holmgren, M., Hughes, T., Kosten, S., van de Leemput, I., Nepstad, D., et al.: Creating a safe operating space for iconic ecosystems, Science, 347, 1317–1319, doi:10.1126/science.aaa3769, 2015.

Schellnhuber, H. J.: Discourse: Earth System Analysis - The Scope of the Challenge, in: Earth System Analysis: Integrating Science for Sustainability, edited by Schellnhuber, H. J. and Wenzel, V., chap. 1, pp. 3–195, Springer, Berlin/Heidelberg, doi:10.1007/978-3-642-52354-0_1, 1998.

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

Schellnhuber, H. J.: Tipping elements in the Earth System, Proceedings of the National Academy of Sciences of the United States of America, 106, 20 561, doi:10.1073/pnas.0911106106, 2009.

Schultz, P., Heitzig, J., and Kurths, J.: Detours around basin stability in power networks, New Journal of Physics, 16, 125 001, doi:10.1088/1367-2630/16/12/125001, 2014.

Ser-Giacomi, E., Rossi, V., López, C., and Hernández-García, E.: Flow networks: A characterization of geophysical fluid transport, Chaos: An Interdisciplinary Journal of Nonlinear Science, 25, 036 404, doi:10.1063/1.4908231, 2015.

Singh, R., Reed, P. M., and Keller, K.: Many-objective robust decision making for managing an ecosystem with a deeply uncertain threshold response, Ecology and Society, 20, 1–32, 2015.

Sontag, E. D.: Mathematical control theory: Deterministic Finite Dimensional Systems, Springer, 2nd edn., 1998.

Steffen, W., Richardson, K., Rockström, J., Cornell, S., Fetzer, I., Bennett, E., Biggs, R., Carpenter, S. R., de Wit, C. a., Folke, C., Mace, G., Persson, L. M., Veerabhadran, R., Reyers, B., and Sörlin, S.: Planetary Boundaries: Guiding human development on a changing planet, Science, p. 1259855, doi:10.1126/science.1259855, 2015.

Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M.: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324, 2013.

Stommel, H.: Thermohaline Convection with Two Stable Regimes of Flow, Tellus, 13, 224–230, doi:10.1111/j.2153-3490.1961.tb00079.x, 1961.

van Kan, A., Jegminat, J., Donges, J. F., and Kurths, J.: Constrained basin stability for studying transient dynamics in complex systems, in review, 2015.

Vaughan, N. E. and Lenton, T. M.: A review of climate geoengineering proposals, Climatic Change, 109, 745–790, doi:10.1007/s10584-011-0027-7, 2011.