Causal attribution in ecosystems with tipping points

Inaugural-Dissertation zur Erlangung der Doktorwürde

Dr. philosophiae

der Fakultät für Umwelt und Natürliche Ressourcen

der Albert-Ludwigs-Universität Freiburg i. Brsg.

vorgelegt von

Michael Stecher

Freiburg im Breisgau

2023

Dekan: Prof. Dr. Heiner Schanz

Betreuer und Referent: Prof. Dr. Stefan Baumgärtner

Zweitbetreuer: Prof. Dr. Christian Möllmann

Koreferent: Prof. Dr. Hermann Held

Datum der Disputation: 16. Mai 2023

Acknowledgments

I would like to thank everyone who has supported me over the past three years and helped to make this dissertation possible. Special thanks to:

- Stefan Baumgärtner, for giving me the opportunity to develop and pursue my own ideas with great liberty, and for giving valuable feedback to become a better researcher.
- Christian Möllmann, for acting as second supervisor and sharing his invaluable expertise on regime shifts in marine ecosystems.
- Hermann Held, for agreeing to examine and read this thesis with keen interest.
- Carsten Dormann, for chairing the disputation and providing stimulating food for thought at the statistics café.
- Philipp Späth, for agreeing to act as examiner and providing a refreshing perspective.
- Martin Quaas and the whole marEEshift consortium, for valuable discussions despite meeting online most of the time.
- Christian Mittelstaedt, Nora Felber and Nadja Leschka, for companionship and mutual support.
- Esther Muschelknautz and Djahane Banoo Salehabadi, for boosting my productivity by offering helpful workshops.
- Clara, for always being there for me, proofreading countless times, and helping me stay motivated all this time.

Summary

How to ascertain causal relationships has been a key question in science and philosophy for centuries. In ecosystems and other complex dynamical systems, determining the causes of a specific system state is particularly difficult. For instance, a fish stock may suddenly collapse after decades of overfishing and progressing climate change. In the presence of tipping points and stochastic influences, it is impossible to know with certainty what has actually caused the collapse. Besides a good understanding of the stock dynamics, systematically attributing an observed system state to its causes thus requires considering probabilistic information. However, there is a lack of adequate concepts and methods for causally attributing the realized or future state of dynamical systems to the varying influence of multiple factors, including agents' deliberate actions, over time.

In this dissertation, I develop conceptual foundations and applied methods for quantifying agents' causal responsibility for the state of dynamical systems, with a focus on ecosystems with tipping points. The goal was to devise a well-founded concept of causal attribution that can be easily operationalized in a wide range of different systems. To achieve this encompassing research goal, I use a variety of methods, including reviewing and synthesizing literature, formalizing abstract ideas, constructing and simulating mathematical models, and calibrating and validating such models with empirical data. The research conducted in this dissertation is divided into three distinct, yet related research papers.

In the first paper, entitled "A stylized model of stochastic ecosystems with alternative stable states", I construct a mathematical model of ecosystems with tipping points that features two different types of stochastic influences: continuous diffusion and discrete jumps. To provide a clear perspective on the subject matter, I review the literature on ecological multistability theory and give precise definitions for its key concepts in the model context. The model thus improves the representation of stochasticity in ecosystems with tipping points and clarifies key concepts of multistability theory. Among other practically relevant applications, the model may be used to determine the probability of regime shift in bistable ecosystems, and how this probability depends on various factors, including management actions.

In the second paper, entitled "Quantifying agents' responsibility: a generalized measure of causation in dynamical systems", I develop a quantitative measure of an agent's causal responsibility for the state of a dynamical system when taking a one-time action. In line with established ideas on causation, I measure the extent to which an agent's action has caused the system state at a later point in time as the degree to which the action is necessary and sufficient for this state. This specification is very general and can be used to attribute the state of a wide range of dynamical systems to human actions and environmental factors. Applying the concept to a number of simple example systems, I find that the extent of causal responsibility crucially depends on the specifics of system dynamics, type of action and the point in time at which the system state occurs.

In the third paper, entitled "Attribution of fish stock collapse to overfishing and climate change", I operationalize causal attribution in a real-world ecosystem, using the recent collapse of the Western Baltic cod stock as a case study. Specifically, I analyze to what extent fishing pressure, climate change and pure chance were causally responsible for tipping the Western Baltic cod stock into a low-productivity regime. I find that the extent to which overfishing has caused the collapse was 75% and climate change 18%. The remaining 7% are attributed to other factors, including stochastic influences. This indicates that unsustainable fishing pressure has been the main driver of the collapse, whereas climate change has altered the stability properties of the stock.

The encompassing concept of model-based causal attribution developed in this dissertation may be used to obtain quantitative knowledge about causal relationships in ecosystems with tipping points and beyond. For instance, the concept allows quantitatively assessing to what extent a realized system state has been caused by different factors, including agents' deliberate actions and pure chance. It may also be used to evaluate an action's effectiveness to reach a given target state as well as its expected causal impact in the future. By quantifying the temporal extent of causal responsibility, the concept provides information about the temporal limits of agents' causal and normative responsibility.

Zusammenfassung

Die Feststellung von kausalen Zusammenhängen ist eine seit Jahrhunderten diskutierte Schlüsselfrage in Wissenschaft und Philosophie. Das Ermitteln der Ursachen eines bestimmten Systemzustands ist besonders schwierig in Ökosystemen und anderen komplexen dynamischen Systemen. Beispielsweise kann ein Fischbestand nach jahrzehntelanger Überfischung und fortschreitendem Klimawandel plötzlich zusammenbrechen. Angesichts von Kipppunkten und stochastischen Einflüssen ist es unmöglich, mit Sicherheit zu wissen, was den Zusammenbruch tatsächlich verursacht hat. Um einen beobachteten Systemzustand systematisch seinen Ursachen zuzurechnen, ist daher neben einem guten Verständnis der Bestandsdynamik die Berücksichtigung probabilistischer Informationen erforderlich. Es mangelt jedoch an geeigneten Konzepten und Methoden zur kausalen Zurechnung des beobachteten oder zukünftigen Zustands dynamischer Systeme auf den zeitveränderlichen Einfluss mehrerer Faktoren, was bewusste Handlungen von Akteuren beinhaltet.

In dieser Dissertation entwickle ich konzeptionelle Grundlagen und angewandte Methoden zur Quantifizierung der kausalen Verantwortung von Akteuren für den Zustand dynamischer Systeme. Ein thematischer Schwerpunkt liegt dabei auf Ökosystemen mit Kipppunkten. Ziel war es, ein gut fundiertes Konzept kausaler Zurechnung zu entwickeln, welches sich in einem breiten Spektrum verschiedener Systeme einfach operationalisieren lässt. Um dieses umfassende Forschungsziel zu erreichen, verwende ich eine Vielzahl von Methoden, darunter der Überblick und die Synthese von Literatur, das Formalisieren abstrakter Ideen, das Entwerfen und Simulieren mathematischer Modelle sowie das Kalibrieren und Validieren von Modellen mit empirischen Daten. Die Forschung in dieser Dissertation ist in drei separate, jedoch thematisch miteinander verbundene Arbeiten gegliedert.

In der ersten Forschungsarbeit mit dem Titel "A stylized model of stochastic ecosystems with alternative stable states" konstruiere ich ein mathematisches Modell von Ökosystemen mit Kipppunkten, das zwei verschiedene Arten von stochastischen Einflüssen aufweist, nämlich: kontinuierliche Diffusion und diskrete Sprünge. Zur verständlichen Darstellung des Themas gebe ich einen Überblick der ökologischen Multistabilitätstheorie sowie präzise Definitionen ihrer wichtigsten Konzepte im Kontext des Modells. Das Modell trägt somit sowohl zu einer besseren Darstellung von stochastischen Einflüssen in Ökosystemen mit Kipppunkten bei, als auch dazu, die Schlüsselkonzepte der Theorie zu schärfen. Neben an-

deren praktisch relevanten Anwendungen kann das Modell dazu verwendet werden, die Kippwahrscheinlichkeit in bistabilen Ökosystemen zu ermitteln und wie diese von verschiedenen Faktoren wie Managementhandlungen abhängt.

In der zweiten Forschungsarbeit mit dem Titel “Quantifying agents’ responsibility: a generalized measure of causation in dynamical systems” entwickle ich ein quantitatives Maß für die kausale Verantwortung eines Akteurs für den Zustand eines dynamischen Systems in Folge einer einmaligen Handlung. In Einklang mit allgemein anerkannten Vorstellungen über Kausalität messe ich den Grad der Verursachung eines Systemzustands durch die vorherige Handlung eines Akteurs als das Ausmaß, zu welchem die Handlung für diesen Zustand notwendig und hinreichend ist. Diese Formulierung ist sehr allgemein und kann dazu benutzt werden, den Zustand eines breiten Spektrums dynamischer Systeme menschlichen Handlungen und Umweltfaktoren zuzurechnen. Bei der Anwendung des Konzepts auf eine Reihe von Beispielsystemen komme ich zum Ergebnis, dass das Ausmaß der kausalen Verantwortung entscheidend abhängt von den Einzelheiten von Systemdynamik und Handlungstyp sowie vom Zeitpunkt, zu dem der Systemzustand eintritt.

In der dritten Forschungsarbeit mit dem Titel “Attribution of fish stock collapse to overfishing and climate change” operationalisiere ich die Zurechnung von Kausalität in einem realen Ökosystem anhand der Fallstudie des kürzlichen Zusammenbruchs des Dorschbestands in der westlichen Ostsee. Konkret analysiere ich, zu welchem Ausmaß Fischereidruck, Klimawandel und Zufallseinflüsse kausal verantwortlich waren für das Kippen des Dorschbestands in ein Regime verminderter Produktivität. Ich komme zu dem Ergebnis, dass Überfischung zu 75% und der Klimawandel zu 18% für den Zusammenbruch verantwortlich waren. Die verbleibenden 7% werden anderen Faktoren wie Zufallseinflüssen zugerechnet. Dieses Ergebnis deutet darauf hin, dass die nicht nachhaltige Fischerei der Haupttreiber des Zusammenbruchs war, während der Klimawandel die Stabilität des Bestands verändert hat.

Das in dieser Dissertation entwickelte umfassende Konzept der modellbasierten Zurechnung von Kausalität kann dazu genutzt werden, quantitatives Wissen über kausale Zusammenhänge in dynamischen Systemen zu gewinnen, beispielsweise in Ökosystemen mit Kipppunkten. Das Konzept ermöglicht es beispielsweise, zu messen, zu welchem Ausmaß ein beobachteter Systemzustand durch verschiedene Faktoren, wie bewusste Handlungen von Akteuren und Zufallseinflüsse, verursacht wurde. Es kann auch dazu verwendet werden, zu beurteilen, wie effektiv eine Handlung ist, um einen bestimmten Zielzustand zu erreichen. Ferner kann es dazu genutzt werden, die erwartete kausale Wirkmacht einer Handlung auf den zukünftigen Systemzustand abzuschätzen. Durch die Quantifizierung des zeitlichen Ausmaßes der kausalen Verantwortung informiert das Konzept zudem über die zeitliche Begrenzung der kausalen und normativen Verantwortung von Akteuren.

Contents

List of Figures

List of Tables

Chapter 1

Synthesis

Many ecosystems are characterized by a duality of apparent stability and gradual or abrupt changes in their state, structure and services over time. Alternating periods of stability, continuous change, and sudden shifts in an ecosystem may arise from the interaction of human actions, external factors, and internal processes. For instance, high fishing pressure might not have a discernible effect on the productivity of a fish stock for years, but may – in combination with other factors, such as climate change – eventually push the stock beyond a tipping point into a low-productivity regime, leading to a potentially irreversible collapse of the stock. Other than fishing pressure, climate change, chance influences, or a combination of these factors could have also played a role in causing the collapse. In such a situation, one fundamental challenge is to determine what has actually caused the collapse. More precisely, one wants to know to what extent each of the factors in question was causally responsible for the regime shift and ensuing collapse of the fish stock.

Quantitative knowledge about the causes of a particular ecosystem state is important for the sustainable management of ecosystems. To evaluate and inform ecosystem management, one would like to attribute a realized system state to some prior management action, and to assess an action's effectiveness to reach a given target state. Due to the complex dynamics, adequate methods for causal attribution in ecosystems with tipping points have so far been lacking. Moreover, existing approaches to measuring causation more generally are deficient in several ways, for instance by not considering dynamics.

In this dissertation, I address this research gap by developing conceptual foundations and applied methods of causally attributing the state of ecosystems with tipping points to agents' actions and other factors. In particular, I generalize the established concept of causal responsibility to dynamical systems with continuously measurable states. I operationalize this concept in ecosystems with tipping points using mathematical models and empirical data to obtain the information required for causal attribution.

The research in this dissertation is structured into three distinct, yet related research papers, which form the Chapters 2-4 of the dissertation. Each paper discusses certain aspects of the overall research focus as well as related aspects and wider ramifications of this research.

In this introductory chapter I summarize and connect the individual papers to each other and to the wider body of literature they are embedded in.

Chapter 1 is organized as follows. In Section 1.1 I introduce the key concepts employed throughout this dissertation and give an overview of the state of the art in the corresponding scientific literature. In Section 1.2 I identify important gaps and open questions in this literature, and outline the research agenda for this dissertation. In Section 1.3 I describe the main results of this research and the methods I used to obtain them. In Section 1.4 I discuss strengths and limitations of my research and sketch potentials for further research. In Section 1.5 I conclude and discuss the relevance of my results.

1.1 Conceptual background

The topic of this dissertation lies at the nexus of multistability theory, stochastic dynamics, causation, and responsibility. That is, I employ the concept of responsibility to measure causation in multistable ecosystems with stochastic dynamics. As a substantial part of the research in this dissertation is conducted at the conceptual level, I discuss each of these concepts in detail.

1.1.1 Multistability theory, tipping points and regime shifts

The fascinating phenomenon of minuscule changes that suddenly have large consequences by triggering self-reinforcing change is common in many natural and human-made systems. The idea reflected in the proverbial “straw that breaks the camel’s back” has come to be known as the “tipping point” following the book of the same title by Malcolm Gladwell (2000). Although usage of the term tipping point has increased strongly, especially in the climate and environmental sciences, it is often not clear which precise mathematical concept it refers to (van Nes et al., 2016). Across various systems and disciplines, tipping points are generally understood as “critical thresholds in a system that, when exceeded, can lead to a significant change in the state of the system” (Hoegh-Guldberg et al., 2018, p. 262).

In ecosystems large, abrupt changes in the system state in response to small disturbances or gradual changes in environmental conditions are commonly known as *regime shifts*. These shifts typically have negative ecological, social and economic consequences and are difficult or impossible to reverse (Scheffer et al., 2001). Regime shifts have been observed in shallow lakes, coral reefs, grasslands, and fisheries (Folke et al., 2004) and may potentially occur in global systems like the Amazon rainforest (Lovejoy & Nobre, 2018), the Antarctic Ice Sheet (Rosier et al., 2021), or the Earth’s climate system (Steffen et al., 2018).

The prevalent theoretical concept to explain how this behavior may arise in ecosystems is the notion of alternative stable states, or: multistability (e.g. Lewontin, 1969; Holling, 1973; Noy-Meir, 1975; May, 1977; Scheffer et al., 2001; Beisner et al., 2003; Petraitis, 2013). Fundamentally, this means that there is more than one locally stable equilibrium state for a certain level of environmental conditions. In the simplest case, there are two locally stable equilibria separated by an unstable intermediate equilibrium. The ecosystem may switch from one state to the other via two elementary mechanisms. The first is when the system state crosses the unstable equilibrium due to some perturbation, which may be regarded as a critical threshold value or tipping point in the system state. The second is when conditions change beyond a bifurcation point, where one of the locally stable equilibria suddenly ceases to exist. This critical level of conditions may also be regarded as a tipping point. That is, for certain levels of environmental conditions, there is only a single, globally stable equilibrium. The related concept of resilience refers to the distance from the critical level – usually in state, but also in conditions (Ludwig et al., 1997) – and represents the amount of disturbance an ecosystem can absorb without changing its basic function, structure, identity, and controls (Gunderson & Holling, 2001; Walker et al., 2004).

Multistability theory has been very influential both in research on, and in policy and management of, ecosystems (Ludwig et al., 1997; Beisner et al., 2003; Walker et al., 2004; Folke, 2006; Lenton et al., 2008; Scheffer, 2009; Barnosky et al., 2012; Dakos et al., 2019). The theory has been communicated by the use of intuitive heuristic graphs that reduce complex ecosystem processes to a mechanistic relationship. This appealing intuitiveness and simplicity has necessarily rendered many of the key concepts fuzzy and has weakened conceptual boundaries. In part, the use of multistability theory as a "boundary object" (Brand & Jax, 2007) has been useful to facilitate the exchange of ideas across disciplinary borders (Strunz, 2012). However, imprecise terminology and lack of conceptual clarity have also led to misunderstandings and confusion about the meaning of concepts like tipping points, thresholds, or resilience.

1.1.2 Stochastic dynamics

Stochasticity is a hallmark of ecosystem dynamics and becomes manifest as seemingly random fluctuations in the ecosystem state over time. The minute causes of these fluctuations are typically not known and thus appear random, but can be grouped into three main sources: (i) demographic stochasticity, which represents random events of individual mortality and reproduction, (ii) environmental stochasticity, which includes variation in climatic conditions and resource availability as well as disturbances like storms, wildfires, floods and diseases,

(iii) stochastic “noise” that arises from measurement errors (Lande et al., 2003). Stochastic perturbations to the ecosystem state that originate from demographic stochasticity or measurement errors occur continuously, but are small in magnitude relative to infrequent environmental disturbances, which can have a large effect on the ecosystem state.

In contrast to deterministic dynamics, under which the ecosystem state at any future point in time can in principle be known with certainty, stochasticity renders ecosystem dynamics fundamentally uncertain. That is, any ecosystem state may occur with some probability and large deviations from the average ecosystem state are possible. In this uncertain environment, the ability of ecosystems to maintain stability and persistence of their functions and characteristics in the face of stochastic perturbations and varying environmental conditions is a key property. In general, the ability of dynamical systems to regulate themselves through negative, self-dampening feedback processes is known as homeostasis (DeAngelis et al., 1986). In the context of multistability theory, ecosystem resilience determines the likelihood of tipping from one alternative stable state to the other due to stochastic perturbations (Gunderson & Holling, 2001; Scheffer & Carpenter, 2003).

Random fluctuations of the ecosystem state over time can be included in dynamic mathematical models of ecosystems by the use of stochastic processes, meaning here temporal sequences of random variables. While there are many different types of stochastic processes, an important distinction is between diffusion and jump processes. At an elementary level diffusion processes describe continuous, incremental change over time, whereas jump processes describe discrete, sudden change or movement at random times. In ecosystems diffusion processes may capture continuous fluctuations arising from demographic stochasticity and measurement errors, whereas jump processes can be used to incorporate rare disturbances that occur at random times. The representation of stochasticity in models of ecosystems with alternative stable states has largely been limited to diffusion processes (e.g. Biggs et al., 2009; Contamin & Ellison, 2009), although jump processes have been used occasionally (e.g. D’Odorico et al., 2006). It is possible to combine both types of stochastic processes in a so-called jump-diffusion process, which has been used in fields as diverse as finance (Merton, 1976), soil hydrology (Daly & Porporato, 2006), or neuroscience (Jahn et al., 2011), but not yet in ecosystems with alternative stable states.

1.1.3 Causation and causal attribution

In ecosystems with stochastic dynamics, it is impossible to know with certainty what has caused a given system state. Likewise, one cannot know with certainty the consequences of an action on the future system state in a stochastic ecosystem. For instance, the collapse of a fish stock may have been caused by fishing pressure, climate change, stochastic influences, or a combination of these factors. The question thus arises to what extent the collapse can be *attributed* to the different factors at play. In general, one would like to measure the degree of causation of a particular system state by some factor when there are several factors that have contributed to causing this state.

The focus of the present analysis lies on the causes of a given effect (or: outcome), rather than on the effects of a given cause (Holland, 1986). The latter perspective is employed by causal inference, which measures causal effects as the difference between two potential outcomes of some response variable: treatment versus control (Haavelmo, 1943; Rubin, 1974; Holland, 1986; Angrist & Pischke, 2009) and is relevant for answering different questions.

One area of research that addresses a question similar to the one in this dissertation is climate attribution science, which asks to what extent an extreme weather event can be attributed to anthropogenic climate change (Allen, 2003; Stott et al., 2004; Otto, 2017). In other words, one wants to measure the degree to which a particular extreme weather event, such as a heatwave or a flood, was caused by climate change – rather than by pure chance. Since the climate system is inherently stochastic, causal attribution needs to be based on probabilistic information (Pfrommer et al., 2019). Specifically, probabilistic climate attribution science analyzes how much more likely an extreme weather event has become due to climate change relative to a *counterfactual* climate that would have prevailed without anthropogenic greenhouse gas emissions (Shepherd, 2016). Since one cannot observe the probability of a particular extreme event occurring in the absence of anthropogenic climate change, model-based simulation of a counterfactual climate is required (Allen et al., 2007). Furthermore, a specific event needs to be defined as a discrete domain of the system state, which may be "to a large extent arbitrary" (Hannart et al., 2016).

Assessing the probabilistic impact of climate change relative to a scenario without human intervention in the climate system reflects a particular, counterfactual and probabilistic conception of causation. Its basic idea may be summarized as a cause being "something that makes a difference, and the difference it makes must be a difference from what would have happened without it" (Lewis, 1973, p. 557). This stands in contrast to purely empiricist notions of causation based on observed regularities, which may be summarized as a cause being "an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second" (Hume, 1748, 60).

These two classic definitions relate to the distinction between necessary and sufficient causation. Hume's definition emphasizes sufficiency and may be restated as a cause being a sufficient condition for an outcome – if the cause occurs, the outcome must occur. Lewis' definition emphasizes necessity and may be restated as a cause being a necessary condition for an outcome – if the outcome occurs, the cause must have occurred before. That is, necessary causation captures how strongly a given outcome depends on a specific cause – rather than on alternative causes – whereas sufficient causation reflects the capacity of a given cause to produce a specific outcome (Pearl, 2009b).

In stochastic systems no single factor can be completely necessary for a realized system state, because any system state may be realized due to pure chance. Hence, probabilistic approaches to causal attribution measure how necessary a factor was for a realized system state by assessing the likelihood of this state in the presence or absence of this factor. Likewise, typically no single factor is completely sufficient for a realized system state when multiple interacting causes have partially contributed to this state. Hence, causal attribution also needs to consider how sufficient a factor was for a realized system state by measuring its relative contribution to this state. This aspect is often neglected in existing approaches to measuring causation.

A number of different measures of causation have been proposed in various contexts (Chockler & Halpern, 2004; Vallentyne, 2008; Braham & van Hees, 2009; Pearl, 2009b; Gleiss & Schemper, 2019; Mittelstaedt & Baumgärtner, 2023). These quantitative approaches can be seen as a small sub-area of the modern literature on causation (e.g. Hume, 1739; Mill, 1843; Wright, 1921; Reichenbach, 1956; Bunge, 1959; Hart & Honoré, 1959; Good, 1961; Mackie, 1965; Lewis, 1973; Pearl, 2009b), which has focused on identifying sets of conditions for an action to be considered a cause of an outcome. That is, causation is often treated as a binary relation rather than as a cardinal measure. The few existing approaches to developing such a cardinal measure are flawed in several respects: (i) most measures are not systematically based on principles of causation, (ii) none consider the dynamic aspect of how causal relationships change over time, and (iii) none are consistent across deterministic and stochastic systems.

1.1.4 Responsibility

Questions of causal attribution have also been discussed under the name “responsibility”, for instance in the context of material flow analysis. One focus of this literature is attributing responsibility for greenhouse gas emissions caused by the production and consumption of goods and services (e.g. Bastianoni et al., 2004; Rodrigues et al., 2006; Lenzen et al., 2007). In particular, this literature is concerned with determining the share of emissions that should be attributed to different agents along the production chain, such as producers and consumers. To this end, different accounting principles are compared in terms of whether they achieve a “fair” share (Ferng, 2003). That is, the measures proposed in the context of material flow analysis are largely ad hoc and not systematically based on established ideas about causation. Calling these measures "responsibility" may reflect common usage of the term, but is not in line with the established concept of responsibility in philosophy. Instead, the measures proposed by these treatments confound descriptive and normative aspects of responsibility.

In contrast to this literature, I build on a well-established and clearly defined concept of responsibility (e.g. Klein, 2005; Duff, 2018; Talbert, 2022). The notion of responsibility I employ in this dissertation may summarized as "the ability to give account to somebody for one's actions, and the possibility to be held accountable for them" (Baumgärtner et al., 2018). Responsibility is a multi-layered concept: following Baumgärtner et al. (2018, Sec. 3.1), I distinguish between three distinct aspects of responsibility.[1] *Causal* responsibility ascribes the consequences of an action to its perpetrator in a purely descriptive manner. *Normative* responsibility is about how one ought to act given some normative framework. *Virtuous* responsibility captures whether an agent actually lives up to her normative responsibility by acting accordingly.

To say that an agent is causally responsible for an outcome goes beyond ascertaining that the agent's action has caused the outcome. In particular, agents can only be causally responsible for an outcome if they can choose freely from a range of alternatives that differ qualitatively in their foreseeable consequences (Bovens, 1998). Beyond these basic requirements for being responsible at all, one important property of responsibility is that its extent may be limited. An agent's causal responsibility is limited by factors beyond the agent's control that hamper her ability to effectuate or avoid a particular outcome, such as chance influences. Similarly, the extent of normative responsibility may be limited due to several reasons (Baumgärtner et al., 2018, Sec. 4.4). One important reason is the agent's limited causal responsibility, in that one can only be obliged to do what one is able to do. In modern ethics, this has become known as the Ought-Implies-Can-Principle (Van Inwagen, 1978; Griffin, 1992). An in-depth discussion of the various conditions, meanings and forms of responsibility is beyond the scope of this dissertation and can be found, for instance, in Baumgärtner et al. (2018).

Causal responsibility has been previously used to measure the degree of causation of an outcome by an agent's action in a semi-formal manner (Vallentyne, 2008). This static and purely probabilistic account of responsibility related to necessary causation has been formalized in a stylized managed ecosystem with two discrete regimes (Baumgärtner, 2020). Sufficient causation and the dynamic aspect of how the extent of causal responsibility changes over time have not been studied so far.

[1] For simplicity, I refer to causal responsibility what Baumgärtner et al. (2018) call "ascriptive responsibility".

1.2 Research questions and contributions

There is a need for concepts and methods to causally attribute the state of ecosystems with tipping points to multiple factors, including agents' deliberate actions and pure chance. The research I conduct in this dissertation to address this encompassing research gap is guided by two overarching questions:

Q1: To what extent has the state of a stochastic dynamical system been caused by an agent's prior action, and to what extent has it been caused by other factors, including pure chance?

Q2: How can model-based causal attribution be operationalized and practically applied in ecosystems with tipping points using empirical data?

In answering these questions, I contribute to the literature by (i) clarifying key concepts of, and improving the representation of stochasticity in, multistability theory, (ii) developing a well-founded, generalized measure of causation in dynamical systems, and (iii) operationalizing causal attribution in the context of ecosystems with tipping points. In the individual papers I break these overarching research questions down into more narrowly defined sub-aspects.

1.2.1 Paper 1

The first paper addresses two major weaknesses of ecological multistability theory. First, the theory lacks a rigorous treatment of stochasticity. While the importance of stochastic influences is often acknowledged implicitly in verbal and graphical representations, these influences are typically not adequately considered in mathematical models of ecosystems with alternative stable states. This is in part due to the second issue, which is conceptual vagueness. Many of the key concepts of multistability have become fuzzy and imprecise as their usage has increased beyond their original scope as descriptive ecological concepts. Hence, the research goal for this paper was to address these conceptual issues by synthesizing the state of the art in the literature and clarifying key concepts using a stylized model. The model helps bringing together different discourses on ecosystems with alternative stable states by synthesizing verbal, graphical and mathematical representations of key concepts of multistability theory. The model can be used for a number of potential applications, such as identifying criteria for sustainable ecosystem management in a stochastic viability framework or determining the probability of a regime shift. By considering different stylized management actions, the model opens new avenues for assessing the management of ecosystems with tipping points and stochastic dynamics.

1.2.2 Paper 2

The second paper is concerned with developing a generalized measure of an agent's causal responsibility for the state of a dynamical system. These conceptual foundations for causally attributing the realized or future system state to some prior action are an important element of causal attribution in ecosystems with tipping points and have so far been missing. The main research questions in this paper are:

Q1: To what extent can the realized state of a dynamical system at a particular point in time be attributed to an agent's prior action?

Q2: What is an action's expected causal impact on the unknown future system state?

Q3: How does the extent of an agent's causal responsibility evolve over time, and how does it depend on the type of system and action?

The resulting measure that answers these questions is based on established principles of causation and improves upon existing measures of causation in several ways. In particular, the measure achieves a full attribution of causality and is consistent across deterministic and stochastic systems for both discrete and continuous conceptions of the system state. A dynamic perspective on measuring causation is relevant for a number of applications where an action's consequences dynamically unfold in a non-trivial way. For instance, the measure can be used to attribute a realized system state to its causes, to assess the effectiveness of management actions for given goals, to design economically efficient liability regulations, and to quantify the temporal limits of normative obligations.

1.2.3 Paper 3

In the third paper I operationalize causal attribution in a real-world ecosystem that has recently crossed a tipping point. Using empirical data, I attribute the recent regime shift and ensuing collapse of the Western Baltic cod stock to overfishing and climate change. Although it has been shown that the combined effect of both factors was responsible for the collapse, their precise individual roles in causing the collapse have remained unclear. This case study is related to a long-standing debate in fisheries science about the causes of stock collapses, which has largely been led with qualitative arguments based on anecdotal evidence. The main research questions in this paper thus refer to both a specific knowledge gap regarding the role of overfishing and climate change in causing the collapse of the Western Baltic cod stock, and a general operationalization of causal attribution in ecosystems with tipping points:

Q1: To what extent can the recent collapse of the Western Baltic cod stock be attributed to overfishing and climate change?

Q2: What are the necessary steps for performing causal attribution in ecosystems with tipping points?

Q3: How much confidence can be placed in the results of causal attribution in the face of uncertainty?

In answering these questions, I contribute to the literature in several ways. First, I give a nuanced and quantitative answer to the debate on the causes of fish stock collapses. Second, I provide a general template of causal attribution in real-world ecosystems with tipping points, using empirical data. Third, I shed light on data and model requirements in the practice of causal attribution.

1.3 Methods and results

Achieving the encompassing research goal of conceiving and operationalizing a novel concept of causal attribution in ecosystems with tipping points requires using a diverse range of methods. The methods I employed in the individual papers include synthesis of existing literature, conceptual development, mathematical modeling, numerical simulations, empirical data analysis and model fitting.

1.3.1 Paper 1

To provide a clear perspective on ecosystems with tipping points, I first review and synthesize the literature on ecological multistability theory. On this basis, I give general and consistent verbal definitions for key concepts of ecological multistability theory and connect them with the simple heuristic graphs that are often used to communicate these concepts.

I then construct a stylized ecosystem model that combines a novel deterministic bistability mechanism with a stochastic jump-diffusion process. The model incorporates the two elementary mechanisms of endogenous regime shifts as well as two different types of stochastic influences – continuous diffusion and discrete jumps. Specifically, the model describes the evolution of a single state variable over time by a stochastic differential equation. The deterministic drift term of this equation specifies the equilibria of the system. Depending on environmental conditions, which are described by an ordinary differential equation, the system has either one globally stable equilibrium or two locally stable and one unstable equilibrium. The stochastic part consists of two terms: a diffusion term representing continuous fluctuations of the system state modeled by a Wiener process, and a jump term representing discrete disturbances of the system state modeled by a compound Poisson process. I then formalize verbal definitions of key concepts of multistability theory by restating them rigorously in the model context.

The simplicity of the model allows deriving an analytical solution for the evolution of the ecosystem state over time as well as closed-form expressions for the expected value and variance of the ecosystem state at any point in time. To be able to simulate sample trajectories of the ecosystem state, I discretize the analytical continuous-time model using the Euler-Maruyama scheme. To include management of ecosystems in the model, I introduce three different types of stylized management actions: (i) directly and instantaneously changing the value of the state variable, (ii) modifying the environmental conditions over time, and (iii) altering the system's susceptibility to stochastic influences. Finally, I sketch a number of potential applications of the model, such as finding economically optimal management strategies.

1.3.2 Paper 2

As a basis for the subsequent analysis, I present a simple and general setup that consists of a single stochastic differential equation with known solution that describes the evolution of the system state over time. This includes deterministic systems when the additive stochastic term is zero. There is a single agent that deliberately takes a one-time action at the initial time 0, which modifies the system dynamics. I distinguish between four different action types: (i) directly modifying the initial value, (ii) modifying the value of an attractor, (iii) changing the rate of convergence to a given attractor, and (iv) changing the volatility of the system state. By directly or indirectly affecting the system state, the action also affects the probability of the system state being in a particular interval at particular time.

I subsequently review established philosophical ideas on causation and discuss a number of principles of causal attribution a quantitative measure of causal responsibility should satisfy. I find that an adequate measure of causation should be based on the difference that an action makes in terms of being necessary and sufficient for a given system state, relative to the counterfactual case of not acting. Further, the causality attributed to the agent and to nature should add up to one, so that the system state is fully and disjointly explained by its causes. Finally, the measure should be applicable to a realized system state from an ex-post perspective as well as to an unknown future system state from an ex-ante perspective.

I then propose a generalized measure of causation that satisfies these criteria. The simplified version of this measure for deterministic systems consists of an action's degree of sufficiency which I take as the relative difference between the realized system state at time t and the counterfactual system state that would have resulted at that time. In this certain environment, both the action and natural dynamics are completely necessary for the realized system state, because it could not have occurred without either of them.

The full measure for stochastic systems also considers an action's degree of necessity, which I take as the relative difference between the realized state's probability due to action and its probability of occurring in the counterfactual case of not acting. Here, the degree of sufficiency is measured as the relative difference between the realized system state and the *expected* system state in the absence of action. The ex-ante measure of causation, which reveals an action's causal efficacy in a representative manner, is the expectation, at the time of action, of the ex-post measure.

I apply this measure in a simulation study of four stylized systems, namely: renewable natural resources under both deterministic and stochastic logistic stock dynamics with and without tipping points. The results I obtain from these simulations and a more encompassing analysis of other systems form the basis of a number of conjectures about the long-run behavior of an agent's ex-ante causal responsibility over time. I find that causal responsibility may either vanish asymptotically over time, or it may converge to a finite, constant level. For systems without thresholds, the qualitative long-run development of causal responsibility is determined by the action type. For systems with thresholds, the development of causal responsibility for some action types also depends on whether the system is above or below the threshold prior to or after the action is taken. Finally, I quantitatively describe the temporal extent of causal responsibility by providing a formal definition of the time period during which an action's causal impact on the system is significant.

1.3.3 Paper 3

Attributing the collapse of the Western Baltic cod stock to overfishing and climate change requires an encompassing procedure of calibrating a suitable model with empirical data, defining counterfactual reference scenarios, simulating these scenarios to obtain the required probabilistic information, and feeding this information into an adequate attribution mechanism.

I use annual time series data from 1970 to 2021 to calibrate a stochastic cusp model (Thom, 1975; Cobb & Watson, 1980; Cobb et al., 1983) to the Western Baltic cod stock. The model considers the effect of two interacting drivers in creating discontinuous regime shift dynamics by modeling the stock size as a cubic function of sea surface temperature and fishing pressure. I estimate the coefficients of this model with maximum likelihood using the package *cusp* (Grasman et al., 2010) within the statistical software *R*.

I use the calibrated model to simulate three counterfactual scenarios, in which either fishing pressure, climate change, or both are absent. In each of these scenarios, I analyze the stability properties of the stock and determine the probability of a shift to low-productivity regime, which facilitated the collapse of the stock. In line with multistability theory, I consider both elementary mechanisms of regime shifts when calculating this probability.

The model suggests that the former high-productivity regime ceased to exist and a critical transition to a low-productivity regime took place in 2007. That is, the regime shift was inevitable for the observed levels of fishing pressure and ocean warming. In the baseline scenario where both factors are absent, the probability of regime shift is merely 6.9%. When applying the generalized measure of causation developed in the second paper, I find that fishing pressure and climate change were jointly causally responsible for the shift to 93.1%. To do so, I treat observed levels of fishing and ocean warming as a joint action by "nature". In this case, the attribution focus lies on how necessary overfishing and climate change were for tipping the stock into a low-productivity regime, because this has been identified as the mechanism underlying the collapse.

Finally, I attribute the collapse to fishing pressure and climate change individually using the attribution mechanism proposed by Mittelstaedt & Baumgärtner (2023). This mechanism can be regarded as an extension of the generalized measure developed in the second paper for multiple agents or factors. However, since the mechanism is purely probabilistic and captures only necessary causation, it thus applies only in the special case of systems with two discrete states. The mechanism measures an individual factor's degree of necessity for a regime shift as the marginal increase in the outcome's probability due to this factor. Since both factors take effect simultaneously in reality, one takes the average probability change due to a factor over all hypothetical sequences when adding factors sequentially. Applying this mechanism, I find that the extent to which overfishing has caused the the collapse of the Western Baltic cod stock was 75%, climate change 18%, and other factors 7%.

1.4 Discussion

The concept of model-based causal attribution I have developed in this dissertation is both general and encompassing. It contains a number of existing approaches as special cases, such as the discrete setting studied by Baumgärtner (2020). As such, its scope is not limited to ecosystems with tipping points, but the concept can be readily applied to a wide range of systems that are affected by human actions, such as fisheries, forests, agricultural systems, public health systems, financial markets, or the macroeconomy. The scope could be further increased by extending and generalizing the concept. In particular, it would be possible to combine two currently separate settings: attributing the continuously measurable system state to a single agent's action and natural dynamics (Paper 2), and attributing the dichotomous state of a bistable system to any number of simultaneously acting agents and other factors (Paper 3). Further, the concept may be extended to study the causal impact of multiple actions taken at different points in time by one or more agents.

Applying the concept to other systems requires good knowledge of the structural causal relationships of those systems formalized in a model with decent predictive power. Hence, to increase the number of systems in which this approach can be applied, more research is needed to obtain and formalize the kind of structural systems knowledge necessary for counterfactual simulations. That is, the concept depends on the ability to make reliable predictions of the system state under conditions outside of observed ranges, which needs to be based on a structural understanding of the system. In addition, the concept rests on the assumption that major sources of stochasticity in the ecosystem dynamics are known and well-represented by the model. That is, the approach depends on reliable probabilistic information and fails when such information is not available. This may be the case if there is Knightian uncertainty (Knight, 1921; Keynes, 1921) regarding potential perturbations (i.e., the potential outcomes are known, but not how they are distributed). For this and deeper forms of uncertainty, the concept of probabilistic model-based causal attribution I develop here is not applicable. In poorly understood systems where the high degree of systems knowledge required for this approach is not available, it is preferable to infer causal relationships directly from data (e.g. Pearl, 2009a; Cunningham, 2021). Inference may be helpful to get an idea about cause and effect in those cases, but cannot be regarded as a substitute to causal attribution in general.

Fundamentally, the scope and extent of causal responsibility depend on the object of responsibility, that is, what one is responsible for (Baumgärtner et al., 2018, Sec. 4.2). In particular, the extent of an agent's causal responsibility for the state of a dynamical system depends on how one specifies the system state. For instance, the discharge of saline water by a mining company into a river may facilitate a bloom of toxic brackish-water algae, which in turn leads to a collapse of the fish population in the river. The mining company may be causally responsible to different degrees for the increased salinity of the river water, the algae bloom, the collapse of the fish population, and potentially the financial losses of fishers. Which of these is a relevant object of causal responsibility cannot be determined in general, but needs to be specified. In fact, agents may be simultaneously causally responsible for multiple objects to different degrees due to a single action. For instance, the mining company may be held responsible both for the collapse of the fish population and for the financial losses of fishers, but to different degrees. The need to specify the object of causal responsibility is a double-edged sword: while it brings versatility, it can lead to problems in applications where unambiguity is required.

For a given object, the quantitative extent of an agent's causal responsibility is subject to uncertainty arising from model and data. In particular, there is uncertainty about the precision of point estimates of model coefficients and the quality of the data due to sampling or measurement errors. The resulting differences in the values of model coefficients can lead to substantial differences in the model output. This uncertainty carries over to the measurement of causal responsibility at multiple points and may compound. Counterfactual simulations are particularly prone to such compounded uncertainty due to potentially large variations in state or conditions in counterfactual scenarios. Likewise, uncertainty about the data used to calibrate the model that may arise from measurement error or sampling error reduces the confidence in the quantitative results of causal attribution. For instance, the biological data for Western Baltic cod used in Paper 3 are the output of a statistical stock assessment model (e.g. Nielsen & Berg, 2014; Aeberhard et al., 2018) based on characteristics of fish from reported catches and scientific trawl surveys. Hence, these data are subject to model uncertainty and represent the mean of a distribution of possible values. One way of addressing and quantifying the effect of model and data uncertainty on the extent of causal responsibility is bootstrapping. In this resampling procedure, causal attribution is repeated a large number of times for random draws from the distributions of input data and model coefficients. The resulting distribution of causal responsibility provides information about the uncertainty of the results and can be used to construct confidence intervals for any confidence level.

With these strengths and limitations in mind, one may think about alternative ways of conceptualizing causal attribution. While the use of stochastic, dynamic models and a probabilistic, counterfactual concept of causation are essential for attributing the state of a stochastic dynamical system to an agent's prior action, the particular specifications of these models and measures of causation may be formulated differently. That is, the stylized model developed in the first paper is one of many possible ways of modeling ecosystems with tipping points. The choice of model should be based on how well it describes the system and on its ability to clearly separate the effect of the factors to which the system state is to be attributed. Likewise, there may be other measures that satisfy the principles of causation I stipulated, but it is unlikely that they will be as simple and general as the measure of causal responsibility developed in the second paper. The causal attribution workflow in the third paper appears to be the most practical among other conceivable operationalizations.

1.5 Conclusion

In this dissertation, I have devised a novel concept of model-based causal attribution in ecosystems with tipping points. In particular, I have developed both conceptual foundations and applied methods for quantifying agents' responsibility for the state of dynamical systems, as well as their operationalization in ecosystems with tipping points.

The core element of the concept is a cardinal measure of an agent's causal responsibility for the state of a dynamical system, given some prior action that modified the system dynamics. The measure incorporates existing knowledge about structural causal relationships in a system to assess the action's degree of necessity and sufficiency for the realized or future system state. While focusing on the role of agency in causing the continuously measurable state of a system with stochastic dynamics, the measure can also be applied to simpler problems that do not involve agency and concern discrete states of a deterministic system. By using clear terminology and expressing diffuse causal knowledge in terms of a single number of causal responsibility, the concept helps improving inter- and transdisciplinary communication.

Further, I have demonstrated how existing causal knowledge can be formalized in a stochastic model and how causal attribution can be operationalized in real-world systems using empirical data. In conclusion, I have devised and applied an integrated process of formalizing causal knowledge in a stochastic model, quantitatively measuring agents' causal responsibility, and implementing the concept using empirical data. I have illustrated this general causal attribution workflow using the particularly complex and important case of ecosystems with tipping points as an example.

The concept can be used to attribute an observed system state to its causes or to assess the expected causal impact of different actions on the future system state. This is relevant for formulating feasible management goals, designing liability regulations, appropriately setting economic incentives, or assessing the effectiveness of management actions and policy measures for given goals. Examples include policies aimed at reaching a predefined system state, such as an inflation target, full employment, a public health target (e.g., vaccination rates), or "good status" of freshwater bodies. When judging whether an agent is to blame or praise for the state of a dynamical system, the concept allows causally attributing the system state to the agent's action and natural dynamics. For example, the concept quantitatively measures to what extent a mining company's discharge of pollutants into a river has caused the subsequent collapse of a fish stock.

The insights gained from applying this concept may provide novel perspectives on unresolved questions in a variety of fields. For instance, attributing the collapse of the Western

Baltic cod stock to fishing pressure and climate change provides, for the first time, a nuanced quantitative answer to a question debated in fisheries science for decades – whether stock collapses are caused by overfishing or climate change (Pershing et al., 2015; Palmer et al., 2016; Swain et al., 2016; Pershing et al., 2016; Brander, 2018; Froese et al., 2022). The result that overfishing and climate change were both necessary for the collapse, but to different degrees, supports the hypothesis that climate change alters the stability patterns of marine ecosystems (Möllmann et al., 2015). This information is crucial for sustainable fisheries management, which needs to adapt to changed stability patterns under climate change (Lindegren & Brander, 2018).

Ultimately, knowledge about cause and effect may remain insufficient for some systems due to fundamental limitations on what can be known. These epistemic limitations need to be considered when designing research agendas. Hence, an important challenge in many scientific fields is to identify what can be known in principle and to devise methods for acquiring this knowledge. The concept of model-based causal attribution developed here may be useful for structuring and communicating this knowledge.

Chapter 2

A stylized model of stochastic ecosystems with alternative stable states

This chapter was written with Stefan Baumgärtner.*

This chapter was published as: Stecher, M. & Baumgärtner, S. (2022b). A stylized model of stochastic ecosystems with alternative stable states. *Natural Resource Modeling*, 35(4), e12345

Abstract:
We construct a generic ecosystem model that features the basic mechanisms of alternative stable states as well as two different stochastic influences. In particular, we use a mean-reverting jump-diffusion process to model the evolution of the ecosystem state over time. We review key concepts of multistability theory and the simple heuristics commonly employed to illustrate them. We then provide mathematical definitions for these concepts in the model context. Our contribution to the literature is twofold: we improve the representation of stochasticity in, and clarify key concepts of, multistability theory. The simplicity of the model enables a number of applications, such as finding economically optimal management strategies, identifying criteria for sustainable ecosystem management in a stochastic viability framework, deriving the probability of a regime shift, or empirically identifying the factors which have caused a specific regime shift.

*Author contributions:
Michael Stecher: conceptualization (equal), formal analysis (lead), methodology (lead), software (lead), visualization (lead), writing – original draft (lead), writing – review & editing (equal).
Stefan Baumgärtner: conceptualization (equal), funding acquisition (lead), methodology (supporting), supervision (lead), writing – review & editing (equal)

2.1 Introduction

Many ecosystems are characterized by a duality of apparent stability and a surprising susceptibility to abrupt changes in the ecosystem's state and services (Petraitis, 2013). These changes, or: regime shifts, often occur in a catastrophic manner and may be difficult or impossible to reverse (Scheffer et al., 2001). Regime shifts have been observed, for example, in shallow lakes, coral reefs, grasslands, and fisheries (Folke et al., 2004) and have been hypothesized for global systems like the Amazon rainforest (Lovejoy & Nobre, 2018), the Antarctic Ice Sheet (Rosier et al., 2021), or the Earth's climate at large (Steffen et al., 2018).

The prevalent concept to explain this behavior is the notion of alternative stable states going back to the seminal works of Lewontin (1969), Holling (1973), Noy-Meir (1975) and May (1977). This means that more than one stable equilibrium state of the ecosystem exists for given environmental conditions. The related concepts of critical thresholds, tipping points and resilience have been particularly influential and have stimulated research across disciplines as well as informed policy and management of ecosystems (Ludwig et al., 1997; Beisner et al., 2003; Walker et al., 2004; Folke, 2006; Lenton et al., 2008; Scheffer, 2009; Barnosky et al., 2012; Dakos et al., 2019). Beyond its sound conceptual core, the success of multistability theory has also been due to the appealing intuitiveness with which it has been propagated. In particular, the use of heuristic devices to reduce complex stochastic interactions and processes in ecosystems to a deterministic, mechanistic relationship makes the theory easy to communicate. The dichotomy between complex reality and simple theory has necessarily rendered many of the key concepts fuzzy and weakened conceptual boundaries. In part, the use of multistability theory as a "boundary object" (Brand & Jax, 2007) has been useful to facilitate the exchange of ideas across disciplinary borders (Strunz, 2012). However, imprecise terminology and lack of conceptual clarity have also led to confusion and have created a divide between researchers with different understandings of concepts like alternative stable states, thresholds, or resilience.

In this paper, we construct a generic ecosystem model that incorporates the key elements of multistability theory as well as two different stochastic influences: continuous diffusion and discrete jumps. While the model is more detailed and complex in its treatment of stochasticity, it is simple enough to provide rigorous definitions and a clear understanding of alternative stable states. It thus helps bringing together different discourses of ecosystems with alternative stable states. In addition, our model easily lends itself to a number of applications, such as finding economically optimal management strategies, identifying criteria for sustainable ecosystem management in a stochastic viability framework, deriving the probability of a regime shift, or empirically identifying the factors which have caused a specific regime shift.

In particular, we use a mean-reverting jump-diffusion process to model the evolution of the ecosystem state over time. Jump-diffusion processes have been used in fields as diverse as finance (Merton, 1976), soil hydrology (Daly & Porporato, 2006), or neuroscience (Jahn et al., 2011), but not yet to capture stochasticity in ecosystems with alternative stable states. In this context, either pure diffusion or pure jump processes have been used. Continuous diffusion has been used to capture natural fluctuations in ecosystems, for example, in models of lake eutrophication (Contamin & Ellison, 2009) or early warning signals for regime shifts (Biggs et al., 2009). Mäler et al. (2007) used a specific diffusion process – an Ornstein-Uhlenbeck process – to model natural groundwater table dynamics. Jump processes have been used to model rare disturbances such as fire in savannahs that may switch between tree-dominated and grassland-dominated states (D'Odorico et al., 2006). Our model combines and enhances these existing approaches: we extend the basic Ornstein-Uhlenbeck model by introducing a novel bistability mechanism for endogenous reversible regime shifts and adding a jump process to allow for infrequent disturbances of the ecosystem state.

The paper is organized as follows. In the next section, we review the key concepts and mechanisms of the theory of alternative stable states in ecology. In Section 2.3, we formalize these concepts in a mathematical ecosystem model and introduce stochastic dynamics. In Section 2.4, we sketch a number of potential applications of the model. In Section 2.5, we discuss our model and conclude.

2.2 Theoretical framework

The prevalent concept to explain how abrupt changes in state variables may arise in response to gradual changes in environmental conditions is the notion of alternative (or: multiple) stable states (e.g. May, 1977; Scheffer et al., 2001; Beisner et al., 2003; Petraitis, 2013). This means that more than one stable equilibrium of the state variables exists for given environmental conditions. Alternative equilibria are stabilized by negative feedbacks that counteract deviations of state variables from stable equilibria (DeAngelis et al., 1986) due to perturbations. The domains in state space in which negative feedbacks cause the state variables to return to the same equilibrium after a perturbation are called *basins of attraction*. The boundary between two basins of attraction is called the *separatrix* or “breakpoint curve” (May, 1977) and contains an unstable equilibrium point of the ecosystem state (Petraitis, 2013). An intuitive way to visualize this is the ball-and-cup heuristic, also called stability landscape. Figure 1 shows such a diagram for the simplest possible case with one state variable and two locally stable equilibria.

The horizontal axis measures the value of the state variable, the vertical axis shows the dynamic potential of the system. The position of the ball in the landscape represents the stability of the ecosystem: the ball always rolls downhill; the force attracting the ball are ecological feedbacks. The shape of the landscape is determined by, and constant for, given environmental conditions. Points where the ball comes to rest are equilibria, valleys are basins of attraction. If the ball is pushed over the ridge by a sufficiently strong perturbation the state variable moves into the other basin of attraction ("basin crossing") where feedbacks induce a convergence to the alternative equilibrium. As a consequence, a potentially large shift in the ecosystem state occurs, where the extent of the shift depends on environmental conditions.

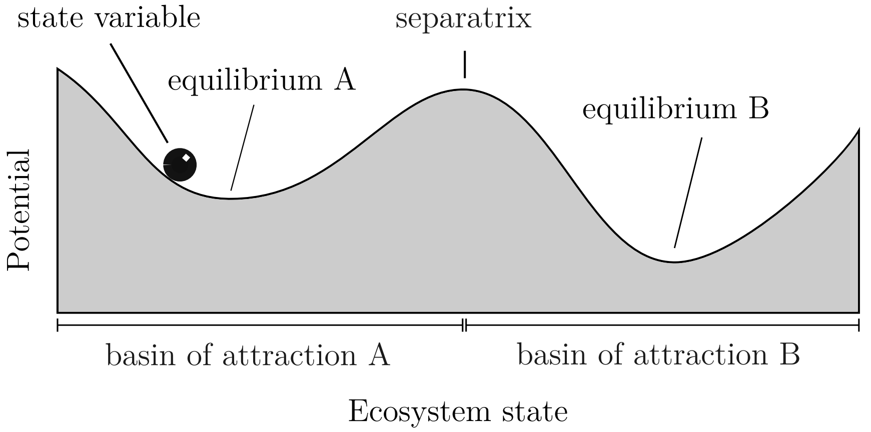


Figure 1: Ball-and-cup diagram

Figure 2 illustrates the effect of changing environmental conditions on the equilibrium ecosystem state. For a low level of conditions only one equilibrium exists at a relatively large value of the state variable. As illustrated in the corresponding stability landscape above, this equilibrium is globally stable, since the ball will always return to the same single valley floor. As conditions increase the stability landscape changes and a second locally stable equilibrium emerges. This enables the possibility of crossing the boundary between alternative basins of attraction due to a perturbation. In this more detailed illustration a second mechanism for abrupt shifts becomes apparent: when conditions change further beyond a level corresponding to point F_2, the first stable equilibrium ceases to exist. If the state variable was attracted by this equilibrium before the feedbacks to the state variable change suddenly, causing an abrupt shift in the ecosystem state ("critical transition"). Reversing environmental conditions to pre-shift levels after a critical transition does not necessarily entail a return of the state variable to pre-shift levels. Ensuring a reverse shift would require changing environmental conditions below a level corresponding to point F_1. The phenomenon that

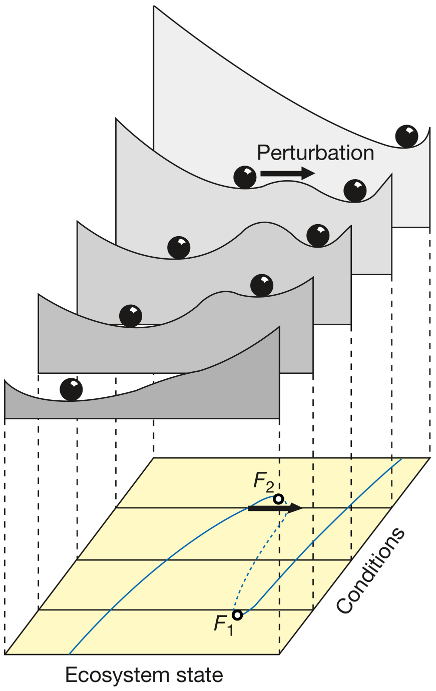


Figure 2: Ball-in-cup diagram and ecosystem response curve: two heuristic devices to illustrate alternative stable states. Reprinted from Scheffer et al. (2001) with permission from Springer Nature.

forward and reverse shifts occur at different critical conditions is known as *hysteresis* and makes critical transitions very difficult to reverse (Scheffer et al., 2001).

Ecosystems exhibit alternative stable equilibria only over a certain range of environmental conditions known as the *bifurcation set* – the range of conditions between the *bifurcation points* F_1 and F_2 in Figure 2. These points mark the location of a fold (or: saddle-node) bifurcation where a single equilibrium bifurcates (or: splits) into three – two locally stable and one unstable – and nonlinear dynamics become possible (Petraitis, 2013). The bifurcation points[2] correspond to critical levels of environmental conditions at which critical transitions between alternative stable states occur. Figure 3 shows the ecosystem response curve in more detail by rotating the bottom plane of Figure 2 clockwise by 90 degrees.

The red arrows represent the effect of ecological feedbacks on the state variable. For constant environmental conditions (i.e., a fixed position on the horizontal axis) the arrows indicate in which direction on the vertical axis the state variable is attracted. The blue curve contains all equilibria of the state variable across a range of conditions. The solid upper and lower branches contain stable equilibria, the dotted section in between represents unstable

[2] Bifurcation points are also referred to as “tipping points”(e.g. Dakos et al., 2019) or “thresholds” (e.g. May, 1977). A discussion of the terminology is given by van Nes et al. (2016). To avoid confusion, we stick to the technical term and use the word threshold only for unstable equilibria between basins of attraction.

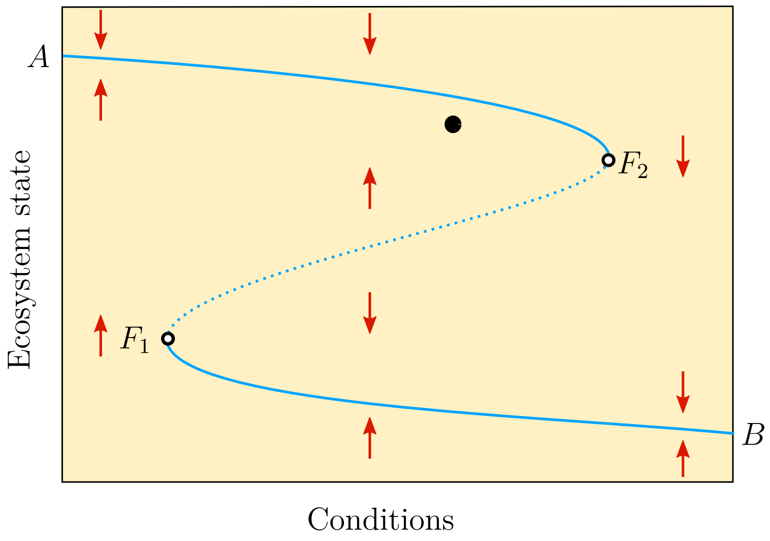


Figure 3: Ecosystem response curve. Redrawn from Scheffer et al. (2001) with permission from Springer Nature.

equilibria on the boundary between the basins of attraction (separatrix). We should distinguish between individual equilibrium points with distinct values of the state variable and collections of equilibrium points with similar, but different values of the state variable. The terms alternative stable states, dynamic regimes and equilibria are often used interchangeably for one or the other concept. We use the terms as follows: for given environmental conditions an *equilibrium* is a unique point on the blue curve with zero rate of change of the state variable and a distinct numerical value attached to it. In contrast, a *dynamic regime* is a set of many equilibrium points and the feedbacks stabilizing them across different environmental conditions – meaning a whole branch of the blue curve and the basins of attraction surrounding it (Scheffer & Carpenter, 2003). Dynamic regimes typically consist of qualitatively similar equilibrium states of the ecosystem with relatively small variation in the equilibrium value of the state variable across a wide range of conditions. For instance, a shallow lake may be in a clear, oligotrophic or a turbid, eutrophic regime.[3] With this, a *regime shift* is defined as a shift from one dynamic regime to the alternative one.

In Figure 3, the vertical distance between the current value of the state variable (indicated by a black dot) and its threshold value (represented by the dotted line) may be interpreted as a measure of resilience (Kinzig et al., 2006).[4] In an elementary sense we understand resilience as a descriptive ecological concept meaning the amount of disturbance an ecosystem

[3] In the clear and turbid regimes, the equilibrium transparency of the lake water changes only slightly across a wide range of nutrient levels – the overall structure and characteristic state of the lake (whether it is clear or turbid), as well as the feedbacks stabilizing it, remain unchanged within a dynamic regime.

[4] A similar definition of the ability to withstand shocks is called *resistance* by Harrison (1979) and Grafton et al. (2019).

can absorb without changing its basic function, structure, identity, and controls (Gunderson & Holling, 2001; Walker et al., 2004). In the particular case of a single state variable we define *resilience* as the maximum possible magnitude of a perturbation of the state variable without entering an alternative basin of attraction. Resilience changes considerably with varying environmental conditions (Carpenter, 2003).[5]

So far, we have discussed the theory of alternative stable states in a deterministic world in which the dynamic behavior of ecosystems is predictable. In reality, ecosystems are subject to stochastic perturbations arising from continuously occurring fluctuations and rare disturbances which cause unexpected and random behavior. In this uncertain world resilience is a key property of ecosystems with alternative stable states, because it determines the likelihood of flipping from one regime to the other (Gunderson & Holling, 2001). In many cases, erosion of resilience by changing environmental conditions makes the shift to an alternative regime due to stochastic perturbations more likely (Scheffer & Carpenter, 2003). We focus on the interaction between stochastic perturbations and the two key deterministic mechanisms for regime shifts (basin crossing and critical transitions) in detail in the next section and leave aside other mechanisms for the occurrence of abrupt shifts in state variables, such as phase shifts (Scheffer et al., 2001).

2.3 Model

We now develop a formal model based on the concepts discussed in the previous section. We first present the deterministic dynamics under constant conditions, before turning to stochasticity and changing environmental conditions. Finally, we introduce management.

2.3.1 Deterministic dynamics, states and regimes of the system

At any point in time $t \in \big[0, \infty\big)$, the state of the ecosystem is characterized by the value of a continuous state variable $X_t \geq 0$, which captures the numerical value of some important quantity in the system, for instance the spawning stock biomass of a fish species or an index of the (multidimensional) ecosystem state. Its evolution over time is given by:

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \theta\big(\mu(c) - X_t\big) + \frac{\mathrm{d}Z_t}{\mathrm{d}t}\,, \tag{1}$$

[5] A minimal mathematical model of the dynamics described in this section can be formulated as $\mathrm{d}x/\,\mathrm{d}t = l - bx + x^k/(x^k + h^k)$, where x is the state variable, l is a factor that promotes x, b and r are the rates at which x decays and recovers, and h is a threshold at which the last term increases steeply, with the steepness determined by k. For the exemplary case of shallow lakes, x are suspended nutrients, l is nutrient loading, b is the nutrient removal rate and r represents internal nutrient recycling (Scheffer et al., 2001).

where $\theta > 0$ parametrizes the strength of feedbacks from ecological processes to the state variable. The parameter $\mu(c)$ determines the equilibrium value of X_t in the absence of stochastic influences and depends on the underlying environmental conditions, which are denoted by the normalized parameter $c \in [0, 1]$. For now, c is constant; we consider changing environmental conditions in Section 2.3.3. Z_t represents stochastic perturbations, such as fluctuations in external forcing or rare events like pest outbreaks. To start with, we discuss the deterministic part of (1) (i.e., the first summand) and elaborate on the stochastic component in Section 2.3.2. That is, we set $Z_t = 0$ for all t. Then, the evolution of the state variable is given by:

$$X_t = X_0 \, \mathrm{e}^{-\theta t} + \mu(c) \left(1 - \mathrm{e}^{-\theta t}\right) . \tag{2}$$

The *deterministic equilibrium* satisfies $\mathrm{d}X_t / \mathrm{d}t = 0$ with $Z_t = 0$ and, from Equation (1), is given by $X_t = \mu(c)$. Thus, the equilibrium ecosystem state is determined by environmental conditions, as proposed by multistability theory (Figure 3). The rate of increase of the state variable is positive when $X_t < \mu(c)$ and negative when $X_t > \mu(c)$. Thus, deterministic equilibria of (1) are stable. The rate at which the state variable converges to its deterministic equilibrium $\mu(c)$ is determined by the parameter θ. The larger θ, the greater the speed of convergence towards the equilibrium value. Equation (1) describes an ecosystem with multiple stable states when $\mu(c)$ takes on more than one possible value for a given level of c across a certain range of environmental conditions. In particular, the bi-stable case depicted in Figure 3 is obtained when there are three possible values of $\mu(c)$ for a given value of c across the range of conditions $F_1 \leq c \leq F_2$ (corresponding to the interval on the horizontal axis between the bifurcations points in Figure 3). In this case, one may rewrite (1) as:

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \begin{cases} \theta(\mu_A(c) - X_t) + \dfrac{\mathrm{d}Z_t}{\mathrm{d}t}, & \begin{array}{l}\text{for } 0 \leq c < F_1 \\ \text{or } F_1 \leq c \leq F_2 \,\wedge\, X_t > \mu_*(c)\end{array} \\[2ex] \dfrac{\mathrm{d}Z_t}{\mathrm{d}t}, & \text{for } F_1 \leq c \leq F_2 \,\wedge\, X_t = \mu_*(c) \\[2ex] \theta(\mu_B(c) - X_t) + \dfrac{\mathrm{d}Z_t}{\mathrm{d}t}, & \begin{array}{l}\text{for } F_1 \leq c \leq F_2 \,\wedge\, X_t < \mu_*(c) \\ \text{or } F_2 < c \leq 1\end{array} \end{cases} , \tag{3}$$

where $\mu_*(c)$ represents unstable equilibria located on the separatrix (corresponding to the dotted blue line in Figure 3) with a corresponding *threshold value* of the state variable that varies with environmental conditions. If $X_t > \mu_*(c)$ the deterministic equilibrium is given by $\mu_A(c)$, and by $\mu_B(c)$ if $X_t < \mu_*(c)$. Together with (1), it follows that $\mu_A(c)$ and $\mu_B(c)$ are locally stable equilibria of X_t. In Figure 3, equilibria with subscript A are points located on the upper branch of the blue curve in Figure 3, those with subscript B on the lower branch. The basins of attraction $b\big[\mu_A(c)\big]$ and $b\big[\mu_B(c)\big]$ comprise the set of all points in state space that converge over time either to $\mu_A(c)$ or to $\mu_B(c)$, respectively, for given environmental conditions:

Definition 1. The *basin of attraction* $b\big[\mu(c)\big]$ is the set of all values of X_t for which

$$\lim_{t\to\infty} X_t = \mu(c)\,, \tag{4}$$

given Equations (1), (3) and $Z_t = 0$ for all t.

To generalize these concepts to different environmental conditions we additionally define the concept of *dynamic regimes* – collections of qualitatively similar equilibrium states of the ecosystem across a range of environmental conditions, such as a clear and a turbid regime across different nutrient levels in a shallow lake. This corresponds to the solid upper and lower branches of the blue curve in Figure 3. The dynamic regimes r_A and r_B encompass the set of all basins of attraction corresponding to equilibria with subscript A or B, respectively, over the entire range of conditions:

$$r_A = \Big\{b\big[\mu_A(c)\big]\Big\}_{c=0}^{F_2}, \qquad r_B = \Big\{b\big[\mu_B(c)\big]\Big\}_{c=F_1}^{1}. \tag{5}$$

With that, a *regime shift* occurs when the state variable moves from one regime into the alternative regime. We assume that the ecosystem is initially in regime r_A. At the time of a regime shift the feedbacks to the state variable change abruptly, but not necessarily the value of the state variable itself. Only over time does X_t converge to the alternative equilibrium $\mu_B(c)$, where θ determines the speed of convergence.

2.3.2 Stochastic dynamics

We now specify the stochastic component Z_t to incorporate continuous diffusion and discrete jumps, and analyze its consequences for the system dynamics. To focus on the stochastic dynamics, we begin with a case in which only one stable equilibrium exists (i.e., $c < F_1$ or $c > F_2$) and regime shifts are not possible. Multiplying (1) by $\mathrm{d}t$ and specifying $\mathrm{d}Z_t = \sigma\,\mathrm{d}W_t + y\,\mathrm{d}N_t$, the evolution of the state variable over time is given by the stochastic differential equation

$$\mathrm{d}X_t = \theta(\mu(c) - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t + y\,\mathrm{d}N_t\,. \tag{6}$$

The right-hand side consists of three additive components: a drift term $\theta(\mu(c) - X_t)$, a diffusion term $\sigma\,\mathrm{d}W_t$, and a jump term $y\,\mathrm{d}N_t$. Hence, Equation (6) describes an Ornstein-

Uhlenbeck (O-U) process[6] (the first two terms) with an additional jump process (the third term). The deterministic drift term, discussed in detail in Section 2.3.1, specifies the change in the expected value of the process over time – the *drift* of the stochastic process X_t (Schuss, 2010).

The diffusion term $\sigma\,\mathrm{d}W_t$ captures continuously occurring perturbations to the state variable, for instance random events of individual mortality and reproduction in population dynamics (Lande et al., 2003). It consists of the diffusion coefficient σ which determines the relative influence of these perturbations on X_t, and the infinitesimal increment $\mathrm{d}W_t$ of a Wiener process. The Wiener process W_t describes Brownian motion: it is a series of identically and independently distributed (i.i.d.) random variables following a normal distribution with zero mean and time-dependent variance. That is, for all $0 \leq s < t$, one has $W_t - W_s \sim \mathcal{N}\big(0, t-s\big)$. The infinitesimal increment $\mathrm{d}W_t = W_{t+\mathrm{d}t} - W_t$ is thus a random variable with mean zero and variance $\mathrm{d}t$.

The jump term $y\,\mathrm{d}N_t$ captures discrete jumps in the value of the state variable, which may arise from rare events like pest outbreaks or extreme weather events and occur at random times. Such behavior can be modeled by a compound Poisson process (Privault, 2013):

$$J_t = \sum_{j=1}^{N_t} y_j \,. \tag{7}$$

The size of jumps is modeled by a random variable y with i.i.d. realizations y_j drawn from a normal distribution with mean $\bar{y}$ and variance β^2. The individual jumps can be observed when they happen, for instance when a hurricane hits a reef and reduces the coral cover. The arrival of jumps follows a homogeneous Poisson counting process N_t with intensity $\lambda > 0$. That is, the probability of n jumps occurring up to time t is given by:

$$P(N_t = n) = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!} \,. \tag{8}$$

Over an infinitesimally small time interval $\mathrm{d}t$, there may be either a single jump or no jump. Hence, the infinitesimal Poisson increment $\mathrm{d}N_t = N_{t+\mathrm{d}t} - N_t$ is drawn from a Poisson distribution with mean $\lambda\,\mathrm{d}t$ (Chiarella et al., 2015):

$$\mathrm{d}N_t = \begin{cases} 1 & \text{with prob.} \quad \lambda\,\mathrm{d}t, \\ 0 & \text{with prob.} \quad 1 - \lambda\,\mathrm{d}t \,. \end{cases} \tag{9}$$

[6] The O-U process was originally introduced by Uhlenbeck & Ornstein (1930) to model the velocity of a Brownian particle.

Thus, a jump occurs with probability $\lambda\,\mathrm{d}t$ in the time interval $\mathrm{d}t$, causing the value of X_t to jump discontinuously by the random amount y_j at the jump time t_j. In between jumps, the state variable follows the O-U process. Table 1 summarizes the parameters of the overall process (6).

Table 1: Parameters describing the stochastic process X_t

	Symbol	Parameter name	Ecological interpretation
Drift term	θ	Mean reversion speed	Strength of ecological feedbacks
	$\mu(c)$	Mean reversion level (of diffusion process)	Deterministic equilibrium value (depending on conditions)
Diffusion term	σ	Diffusion coefficient	Strength of random fluctuations
Jump term	y	Jump size (random variable)	Magnitude of rare events
	$\bar{y}$	Mean jump size	Average magnitude of rare events
	β^2	Variance of jump size	Variability of rare events
	λ	Intensity of Poisson process	Frequency of rare events

The stochastic differential Equation (6) describes the evolution of X_t over the infinitesimally small time interval $\mathrm{d}t$. To obtain the evolution of X_t over the entire time interval $[0, \infty)$, we solve (6) with the initial condition $X_{t=0} = X_0 > \mu_*(c)$, given $\mathrm{d}W_t$ as the infinitesimal increment of a Wiener process and $\mathrm{d}N_t$ according to (7), (8) and (9). Assuming for the moment that no regime shifts occur (for instance, $c < F_1$ or $c > F_2$), this initial value problem has the general solution (Appendix A):

$$X_t = X_0\,\mathrm{e}^{-\theta t} + \mu(c)\left(1 - \mathrm{e}^{-\theta t}\right) + \sigma \int_0^t \mathrm{e}^{-\theta(t-s)}\,\mathrm{d}W_s + \int_0^t \mathrm{e}^{-\theta(t-s)} y_s\,\mathrm{d}N_s\,. \tag{10}$$

At time $t = 0$, X_t is equal to the observable initial value X_0. For any later point in time, the deterministic part of (10) can be calculated. For the stochastic part, the realizations of the Wiener and compound Poisson process are not known ex-ante, but one can calculate their expected value. The Wiener process has an expected value of zero by definition as its increments are drawn from a normal distribution with zero mean. For the jump term, the expected value with respect to the frequency and size of jumps is given by:

$$\mathbb{E}_{y,\mathrm{d}N}\left[\int_0^t \mathrm{e}^{-\theta(t-s)} y_s\,\mathrm{d}N_s\right] = \bar{y}\,\frac{\lambda}{\theta}\left(1 - \mathrm{e}^{-\theta t}\right)\,. \tag{11}$$

The expected value of the jump component of (10) consists of the expected size $\bar{y}$ and the arrival rate λ of jumps. The absolute value of this expression is decreasing in θ (proof in Appendix A), which means that the relative contribution of jumps to the expected value of the state variable depends negatively on the strength of deterministic feedbacks. The expected value of X_t is thus given by:

$$\mathbb{E}[X_t] = X_0\,\mathrm{e}^{-\theta t} + \mu(c)\left(1 - \mathrm{e}^{-\theta t}\right) + \bar{y}\,\frac{\lambda}{\theta}\left(1 - \mathrm{e}^{-\theta t}\right)\,, \tag{12}$$

and its variance is given by (Das, 2002):

$$\mathrm{Var}[X_t] = \frac{\sigma^2 + \lambda\beta^2}{2\theta}\left(1 - \mathrm{e}^{-2\theta t}\right) + \mathbb{E}[X_t]^2\,. \tag{13}$$

Over time, the expected value of X_t tends away from its initial value X_0 and towards $\mu(c)$ as a result of deterministic ecological feedbacks, but is perturbed by random jumps. In the limit, X_t converges to its stationary mean, which is known as *mean reversion*:

$$\lim_{t\to\infty} \mathbb{E}[X_t] = \mu(c) + \bar{y}\,\frac{\lambda}{\theta}\,. \tag{14}$$

That is, the state variable is expected to converge to its deterministic equilibrium $\mu(c)$ plus a deviation due to jumps. The expected deviation from the equilibrium due to rare events depends positively on the arrival rate λ and the mean size of jumps $\bar{y}$, and negatively on the strength of deterministic feedbacks θ, which counteract the effect of random jumps. These results only hold when a single value of $\mu(c)$ exists for a given level of c (i.e., $c < F_1$ or $c > F_2$) and regime shifts are not possible. When $\mu(c)$ may take on three values (i.e., $F_1 \leq c \leq F_2$), stochastic perturbations can induce endogenous regime shifts and we cannot make closed-form statements about the behavior of X_t over an infinite time horizon.

Consider the case of a single regime shift at time t_{RS} due to basin crossing under constant environmental conditions. At this point in time X_t falls below its threshold value $\mu_*(c)$ and the state variable moves from regime r_A into the alternative regime r_B. Once in the alternative regime, the feedbacks acting on the state variable change instantaneously and X_t is attracted by its alternative deterministic equilibrium $\mu_B(c)$. The evolution of X_t after the shift is described by the same stochastic process (10) as before, but resets at time t_{RS} with initial value $X_{t_{\mathrm{RS}}}$ and the alternative equilibrium $\mu_B(c)$. This is possible because the process X_t (Equation 6) fulfills the *Markov property*: future values of X_t depend solely on the current value of the process and not on past realizations – the process is memoryless. In case of a single regime shift at time t_{RS}, one can rewrite (10) more precisely as:

$$X_t = \begin{cases} X_0\,\mathrm{e}^{-\theta t} + \mu_A(c)\left(1-\mathrm{e}^{-\theta t}\right) + \sigma\int_0^t \mathrm{e}^{-\theta(t-s)}\,\mathrm{d}W_s + \sum_{j=1}^{N_t}\mathrm{e}^{-\theta(t-t_j)}y_j & \text{for } 0 \le t < t_{\mathrm{RS}} \\ X_{t_{\mathrm{RS}}}\mathrm{e}^{-\theta(t-t_{\mathrm{RS}})} + \mu_B(c)\big(1-\mathrm{e}^{-\theta(t-t_{\mathrm{RS}})}\big) + \sigma\int_{t_{\mathrm{RS}}}^t \mathrm{e}^{-\theta(t-s)}\,\mathrm{d}W_s + \int_{t_{\mathrm{RS}}}^t \mathrm{e}^{-\theta(t-s)}\,y_s\,\mathrm{d}N_s\,, & \\ & \text{for } t \ge t_{\mathrm{RS}} \end{cases} \tag{15}$$

where N_t is the number of jumps that have occurred up to time t and t_j is the time of jump j. At time t_{RS}, one has observed the times and sizes of all jumps up to this point and the first line of Equation (15) gives the value of X_t at every prior time from an ex-post perspective. The further development beyond t_{RS} is not known ex-ante: one can calculate the expected value based on the updated initial value of the state variable $X_{t_{\mathrm{RS}}}$, which is known with certainty. Equation (15) holds until the next regime shift happens, say at t_{RS2}, when the process resets again. This succession of regime-shift-and-resetting can go on indefinitely, but updating (15) every time a regime shift occurs accurately describes the dynamics.

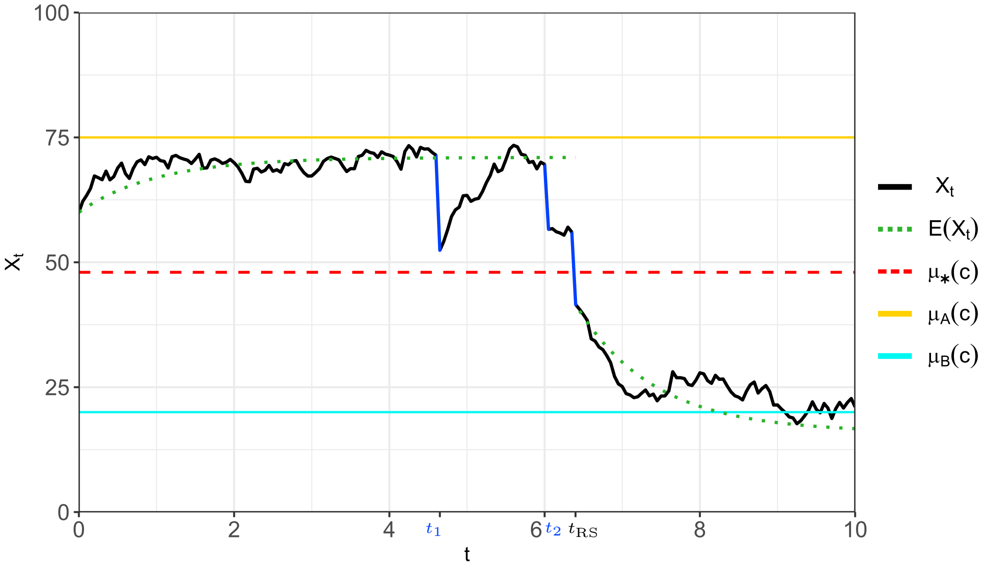


Figure 4: Sample path for a random realization of the stochastic process X_t with a forward regime shift. Parameter values: $X_0 = 60, \mu_A(c) = 75, \mu_B(c) = 20, \mu_*(c) = 48, \theta = 1, \sigma = 5, \lambda = 0.4, \bar{y} = -10, \beta = 5$. The simulation was performed with the Euler-Maruyama discretization scheme using time steps of $\Delta t = 0.05$.

Figure 4 depicts the case of a regime shift due to basin crossing caused by a negative jump. In this realization of the stochastic process (15), the state variable is initially below its deterministic equilibrium $\mu_A(c)$ by which it is attracted continuously over time. Stochastic diffusion causes the state variable to fluctuate and thereby keeping it from actually reaching the equilibrium. The first jump at t_1 brings the state variable precariously close to its threshold value, but the system is able to recover from this perturbation due to ecological feedbacks. The second jump at t_2 is smaller, but stochastic diffusion counteracts the deterministic feedbacks. The next jump happens before the system can recover and pushes the state variable below its threshold value, causing a shift to the alternative regime r_B at time t_{RS}. That is, the ecosystem is resilient against the first two jumps, but cannot cope with the additional perturbation of another negative jump in its state of decreased resilience. When the stochastic process resets at time t_{RS} the expected value of the further development is formed anew. Over time, the new regime stabilizes itself as the state variable is attracted by its new deterministic equilibrium $\mu_B(c)$; resilience against a reverse shift back to the initial regime r_A increases.

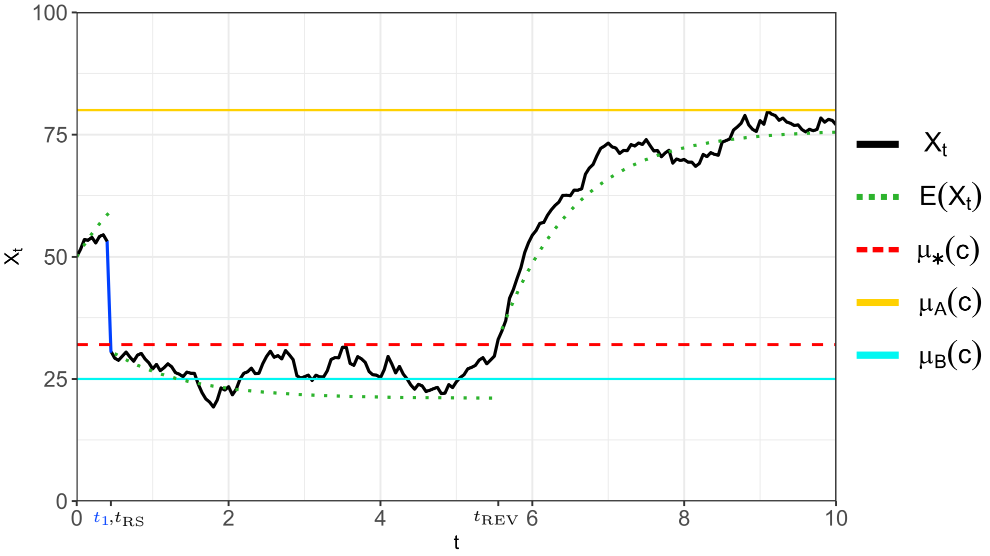


Figure 5: Sample path for a random realization of the stochastic process X_t with a reverse regime shift. Parameter values: $X_0 = 50, \mu_*(c) = 32, \mu_A(c) = 80, \mu_B(c) = 25, \theta = 1, \sigma = 5, \lambda = 0.4, \bar{y} = -10, \beta = 10$. Again, $\Delta t = 0.05$.

Figure 5 shows a situation in which the state variable does not remain in regime r_B after crossing the threshold. After the shift the state variable fluctuates around its deterministic equilibrium $\mu_B(c)$ which is located close to the threshold. This represents a case where conditions are unfavorable for regime r_B (i.e., c is only slightly greater than F_1, compare Figure 3) and resilience against a shift to regime r_A is low. Stochastic diffusion causes the state variable to cross the threshold between the basins of attraction a second time at t_{REV}. After this reverse regime shift, the state variable quickly converges to its deterministic equilibrium $\mu_A(c)$ due to ecological feedbacks.

2.3.3 Changing environmental conditions

So far, we have focused on ecosystem dynamics under constant environmental conditions. In reality, "conditions are never constant" (Scheffer et al., 2001). That is, we have

$$c = c(t) \ \text{ with } \ c(0) = c_0\,, \tag{16}$$

which influences the dynamics of the state variable by changing the deterministic equilibrium $\mu(c)$ as well as the threshold value $\mu_*(c)$ continuously over time. For simplicity, we assume that changes in environmental conditions are deterministic and thus foreseeable.

Essentially, conditions are quantities in the ecosystem that change very slowly relative to state variables (Beisner et al., 2003).[7] A useful special case of (16), which we will assume in the following, is the basic exponential convergence process:

$$c_t = c_0 + \Delta c\,(1 - \mathrm{e}^{-\gamma t})\,, \tag{17}$$

where Δc indicates the absolute change in normalized conditions c and γ parametrizes the rate of convergence. We assume that $0 < \gamma \ll \theta$, that is, environmental conditions change much less quickly than the state variable. When environmental conditions change, this modifies the equilibria of the system and the values of $\mu_A(c), \mu_B(c)$ and $\mu_*(c)$ change. That is, changing environmental conditions have no instantaneous effect on the value of state variable, but influence its deterministic trend over time and its susceptibility to regime shifts. Taken together, changing environmental conditions (Equation 17) and state dynamics (Equation 10) result in a continuously ongoing dual adjustment process.[8] Figure 6 illustrates the resulting dynamics.

As described in Section 2.2, changing environmental conditions pose an additional mechanism for regime shifts. If environmental conditions move beyond one of the bifurcation points, a critical transition to the alternative regime is inevitable, regardless of the value of the state variable. The mechanism for critical transitions is simple: when c increases beyond F_2, equilibrium $\mu_A(c)$ ceases to exist according to (3) and the state variable is attracted by the alternative equilibrium $\mu_B(c)$. Already before environmental conditions actually increase beyond F_2, a regime shift is likely to happen due to stochastic perturbations as a result of

[7] In fact, the rate of change of environmental conditions can be several orders of magnitude slower than the rate of change of state variables (Rinaldi & Scheffer, 2000). For instance, even though the current rate of accumulation of greenhouse gases in the atmosphere is unprecedented in geological history, the resulting changes in climatic conditions unfold relatively slowly compared to the changes in population densities or species abundances they entail.

[8] Formally, in Equations (6), (10), (12) and (15), c is time-dependent according to (17); and in Equation (14) $\mu(c)$ is replaced by $\mu(c_0 + \Delta c)$.

decreased resilience. The two mechanisms for regime shifts often act in combination. In general, a change in environmental conditions influences the resilience of the ecosystem to stochastic perturbations, which determines the likelihood of a regime shift (Gunderson & Holling, 2001, p. 50).

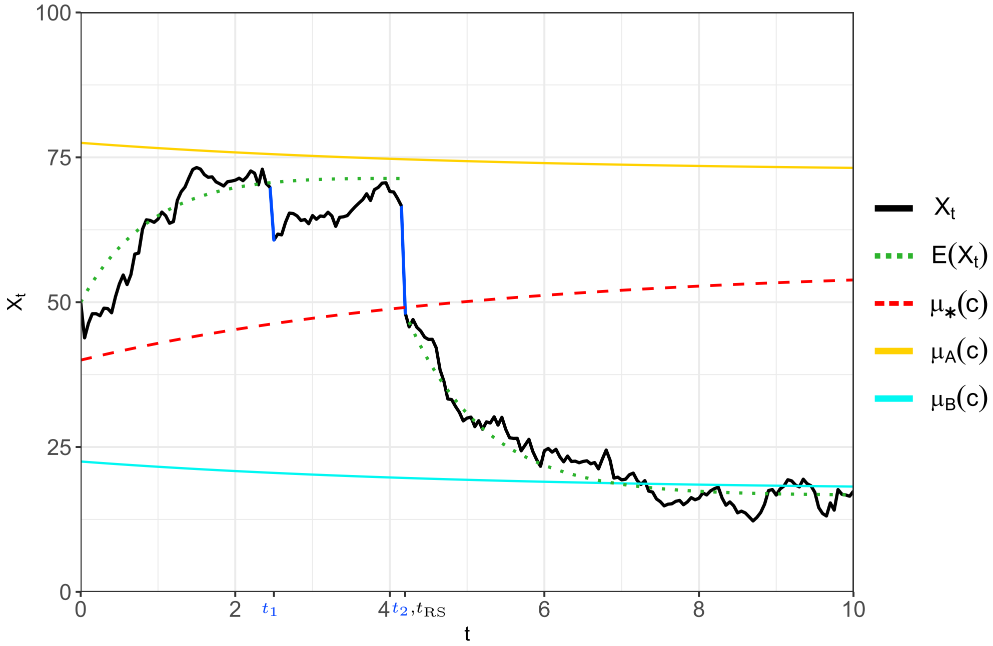


Figure 6: Sample path for a random realization of the stochastic process X_t with changing environmental conditions. Parameter values: $X_0 = 50, \mu_A(c_t) = 90 - 25c_t, \mu_B(c_t) = 35 - 25c_t, \mu_*(c_t) = 80c_t, \theta = 1, \sigma = 5, \lambda = 0.3, \bar{y} = -10, \beta = 5, c_0 = 0.5, \Delta c = 0.2, \gamma = 0.2$. Again, $\Delta t = 0.05$.

Once the state variable is in regime r_B, conditions need to be reversed to less than F_1 to ensure a reverse shift to regime r_A. Hence, our model captures hysteresis of the ecosystem state in response to changing environmental conditions.

2.3.4 Ecosystem management

We include ecosystem management in the model as follows. There is a single ecosystem manager who chooses the type and intensity of a management action $a = \{v, q, z\}$ taken at time $t = 0$. There are three different types of management actions, each of which affects the ecosystem in a different way: type v directly and instantaneously influences the state variable, type q changes the environmental conditions over time, and type z modifies the system's susceptibility to stochastic influences.

Action $v > -X_0$ instantaneously changes the value of the state variable by the amount v at the time of action $t = 0$. Management may increase or decrease the value of the state variable. For instance, if the state variable is the biomass of a fish stock, harvesting a certain amount of fish immediately reduces the state variable by this amount while restocking increases it immediately. Other aspects of the stock dynamics, such as the equilibrium level of the state variable $\mu(c)$ or the threshold level $\mu_*(c)$ are unaffected by this type of management action. If no regime shift occurs the state variable tends to return to its equilibrium level over time due to ecological feedbacks. In this case, the time path of X_t resulting from taking management action v at time $t = 0$ is given by:

$$X_t(v) = (X_0 + v)\, \mathrm{e}^{-\theta t} + \mu_A(c_t)\left(1 - \mathrm{e}^{-\theta t}\right) + \sigma \int_0^t \mathrm{e}^{-\theta(t-s)}\, \mathrm{d}W_s + \int_0^t \mathrm{e}^{-\theta(t-s)} y_s \, \mathrm{d}N_s\,. \tag{18}$$

A shift to regime r_B may occur at any time in analogy to (15) and can be made either more or less likely by management action v. Indeed, for sufficiently strong actions, that is, $X_0 + v < \mu_*$, the state variable falls below its threshold value directly at the time of action $t = 0$ and a regime shift occurs with certainty.

Action $q \in [-c_0 - \Delta c, 1 - c_0 - \Delta c]$ changes the conditions over time by adding the amount q to the exogenous change in conditions Δc according to (17). Again, management may increase or decrease the conditions such that c_t lies in the normalized range $[0, 1]$. This type of management thereby modifies the deterministic equilibrium value $\mu(c)$ and the threshold value $\mu_*(c)$. In contrast to action v, action q does not change the value of the state variable directly, but influences its dynamics by changing the feedbacks acting on the state variable. Since conditions change only slowly relative to the state variable, actions of type q take a longer time to have the same quantitative effect on the state variable than actions of type v. In the example of fish in a lake, suppose there is anthropogenic nutrient loading of the lake, leading to an increase in resource availability for planktivorous fish. The higher availability of feed increases the spawning rates, which increases the equilibrium biomass $\mu(c)$ of planktivorous fish (assuming that death rates remain constant). Due to the Allee effect, increased resource availability may also result in a lower extinction threshold μ_* for the fish stock (Petraitis, 2013, Chap. 2.2). As conditions change over time to their new level $c_0 + \Delta c + q$ with rate γ, the state variable X_t adjusts incrementally to the modified equilibrium value $\mu_A(c_t)$ with rate θ (if no regime shift occurs):

$$X_t(q) = X_0\, \mathrm{e}^{-\theta t} + \mu_A(c_t(q))\left(1 - \mathrm{e}^{-\theta t}\right) + \sigma \int_0^t \mathrm{e}^{-\theta(t-s)}\, \mathrm{d}W_s + \int_0^t \mathrm{e}^{-\theta(t-s)} y_s \, \mathrm{d}N_s\,. \tag{19}$$

Action z modifies one or more of the stochastic parameters $\sigma, \bar{y}, \beta, \lambda$ by the amount z and is bounded by non-negativity constraints for σ, β and λ. With this action, management can modify the susceptibility of the state variable to stochastic perturbations. Examples would be dikes against floods or irrigation systems and water pumps against droughts. The modified time path of X_t is given by:

$$X_t(z) = X_0\, \mathrm{e}^{-\theta t} + \mu_A(c_t)\left(1 - \mathrm{e}^{-\theta t}\right) + \sigma(z) \int_0^t \mathrm{e}^{-\theta(t-s)}\ \mathrm{d}W_s + \int_0^t \mathrm{e}^{-\theta(t-s)} y_s(z)\, \mathrm{d}N_s(z)\,. \quad (20)$$

2.4 Potential applications

Due to its simplicity and generality, the model hands itself to a variety of applications useful for ecosystem management.

2.4.1 Model calibration

Calibrating the model with empirical data makes it possible to understand which processes and factors play an important role in determining the ecosystem state. For the calibration, time series data of a characteristic state variable (or an index of the ecosystem state) and of important environmental conditions is required. After normalizing the conditions c_t to the interval $[0, 1]$, it is possible to fit the parameters of Equations (3) and (6) as well the functional relationship $\mu(c)$ using maximum likelihood estimation. If the data exhibit abrupt regime shifts, knowledge about the threshold value μ_* across different conditions is required, which may be difficult to obtain in practice. In this case, it may be necessary to run an auxiliary model that includes higher power terms of X_t to identify all possible stable and unstable equilibria.

Once calibrated, the model may help in determining the relative importance of different factors (external driver, management action, random variation, rare event) that caused a regime shift. In a further step, one can quantify the extent to which different factors are responsible for a regime shift using the concept of partial responsibility (Vallentyne, 2008; Baumgärtner, 2020). We derive the probabilistic information required for this method in Section 2.4.4.

2.4.2 Optimal management

Suppose the ecosystem manager faces the problem of maximizing expected intertemporal welfare derived from net benefits enjoyed from the ecosystem. These benefits, denoted by $\pi(r, a)$, depend on the chosen management action and differ between regimes. They consist of different levels of ecosystem services or direct economic benefits, such as harvest. Specifically, assume that the manager receives a flow of benefits $\pi_t(r_A, a) \neq \pi_t(r_B, a)$, irrespective of the precise level of the state variable X_t. Since the dynamic regimes are ultimately defined by the value of the state variable by (3), (5) and Definition 1, we rewrite the benefits as $\pi_t(X_t, a)$.

The manager must choose a single management action a of type v, q or z at $t = 0$. She can choose from all feasible management actions described in Section 2.3.4, but incurs costs of $\kappa_t(a)$ associated with the action. We make no assumptions on the shape or time profile of $\kappa_t(a)$, other than it being a convex function. Social welfare is measured using a well-behaved utility function $U(\cdot)$, that is, $U'(\cdot) > 0, U''(\cdot) < 0$, and a time preference rate ρ. Hence, the manager needs to solve the problem

$$\max_a \mathbb{E}\left[\int_0^\infty \mathrm{e}^{-\rho t}\, U[\pi_t(X_t(a), a) - \kappa(a)]\,\mathrm{d}t\right] \tag{21}$$

subject to (15), (17),

$$\pi_t(X_t, a) = \begin{cases} \pi_t(r_A, a) & \text{for } X_t \geq \mu_* \\ \pi_t(r_B, a) & \text{for } X_t < \mu_* \,, \end{cases} \tag{22}$$

and

$$X(0) = X_0(a);\ 0 \leq X_t \leq 1; \qquad c(0) = c_0;\ 0 \leq c_t \leq 1\,. \tag{23}$$

This problem cannot, in general, be solved analytically, but can be solved numerically (Kushner & Dupuis, 2001). In particular, the stochastic nature of the model dynamics suggests using dynamic programming techniques suited to deriving optimal feedback control rules rather than open-loop controls to account for uncertain system states (Bellman, 1966). This is an interesting decision problem with two trade-offs: the manager needs to choose not only the optimal intensity of the management action given costs and social risk and time preferences, but also the type of management action. In particular, there is an interesting choice along the temporal dimension between influencing the state variables or the conditions in the model (setting the option of management action z influencing the stochastic parameters momentarily aside). There is a trade-off between an immediate, but relatively short-lived intervention and a slow, persistent change. We would expect that the main factor influencing this decision is the size of the discount rate ρ. Larger values of ρ indicate a stronger time preference for the present and would imply taking management action v. Management action z will be optimal if the manager is very risk-averse.

2.4.3 Viability management

Welfare-maximizing management based on discounted expected utility may not necessarily be sustainable in the sense that long-run costs and benefits tend to be neglected due to utility discounting (De Lara et al., 2015). In addition, economic analyses typically assume good substitutability between between natural and other forms of capital. This notion of weak sustainability (Neumayer, 2003) has been criticized for its inability to cope with multistability and other issues (van den Bergh, 2014). For these reasons, it may be preferable to use evaluation concepts that ensure strong sustainability under conditions of uncertainty and multistability, such as stochastic viability (Béné & Doyen, 2018; Oubraham & Zaccour, 2018; Doyen et al., 2019). The basic idea of stochastic viability is that the continued existence of certain ecosystem functions and components is guaranteed at all times with a sufficient probability (Baumgärtner & Quaas, 2009).

Under the stochastic viability approach an ecosystem manager needs to choose a management action from the set of *viable actions* a^{viab} which consists of those actions that are both admissible (a^{ad}) and that satisfy the state constraint of being above the threshold $X_t \geq \mu_*$ with at least the probability α:

$$a_\alpha^{\text{viab}}(X_0, t=0) = \left\{ a \in a^{\text{ad}} \,|\, P(X_t \geq \mu_*) \geq \alpha \text{ for all } t \right\}, \tag{24}$$

given the uncertain dynamics (15), (17). The solution of this stochastic viability problem can be obtained with dynamic programming methods (Doyen & De Lara, 2010) that can readily be applied to our model.

2.4.4 Probability of regime shift

The probability of flipping into an alternative regime is determined by the state variable's resilience to stochastic perturbations, which in our model is equivalent to the distance of the state variable X_t from its threshold value $\mu_*(c)$. The larger the resilience, the lower is the probability of a regime shift. For a known value of X_t at time t, we can calculate the *instantaneous* probability of a regime shift from r_A to r_B as the probability of the state variable X_t falling below its threshold value $\mu_*(c)$ within the next infinitesimal time interval $\mathrm{d}t$:

$$\begin{aligned} P_t(r_A \to r_B \,|\, X_t) = {} & P\Big(\sigma \,\mathrm{d}W_t + y \leq -\big[X_t + \theta(\mu(c_{t+\mathrm{d}t}) - X_t) - \mu_*(c_{t+\mathrm{d}t})\big]\Big) \cdot \lambda \,\mathrm{d}t \\ & + P\Big(\sigma \,\mathrm{d}W_t \leq -\big[X_t + \theta(\mu(c_{t+\mathrm{d}t}) - X_t) - \mu_*(c_{t+\mathrm{d}t})\big]\Big) \cdot (1 - \lambda \,\mathrm{d}t), \end{aligned} \tag{25}$$

which explicitly considers the two possible cases of either a jump of random size y or no jump occurring. Since we have assumed independence of the three random variables, it is possible to use a single probability distribution for the sum $\sigma \,\mathrm{d}W_t + y \sim \mathcal{N}\big(\bar{y}, \beta^2 + \sigma^2 \,\mathrm{d}t\big)$.

This way of obtaining the probability of flipping into an alternative regime requires knowledge of the specific realization of the stochastic process X_t, which is known only once it has happened, or: *ex-post*. In practice, today's management actions often affect the state of the system in the future and one needs to assess the probabilistic consequences of different actions before taking them, or: *ex-ante*. In this case, one only knows the value of the state variable X_0 at time $t = 0$ and must form expectations about the state of the system at future points in time. In this case, it is possible to use the expected value and variance given in Equations (12) and (13) to calculate the expected instantaneous probability of regime shift at time t, assuming that no shifts have happened until that point in time.

There is a very simple, well-performing approximation of the expected probability of regime shift that is useful for management applications. The probability of a shift from regime r_A to regime r_B taking place at time t, conditional on having taken management action v at time 0 and no shifts having occurred until t, is approximately given by:

$$P_t\big(r_A \to r_B \mid \mathbb{E}[X_t(v)]\big) \approx p_t(v) = \bar{p}_t + \Delta p(v) \cdot \mathrm{e}^{-\theta t}\,, \tag{26}$$

where $\bar{p} = P_t\big(r_A \to r_B \mid \mathbb{E}[X_t]\big)$ indicates the expected baseline probability of regime shift in the absence of management actions. The maximum change in probability due to the management action is denoted by $\Delta p(v)$ and needs to be calibrated. The approximation for management type q is very similar and given by:

$$P_t\big(r_A \to r_B \mid \mathbb{E}[X_t(q)]\big) \approx p_t(q) = \bar{p}_t + \Delta p(q) \cdot \big(1 - \mathrm{e}^{-\gamma t}\big)\ . \tag{27}$$

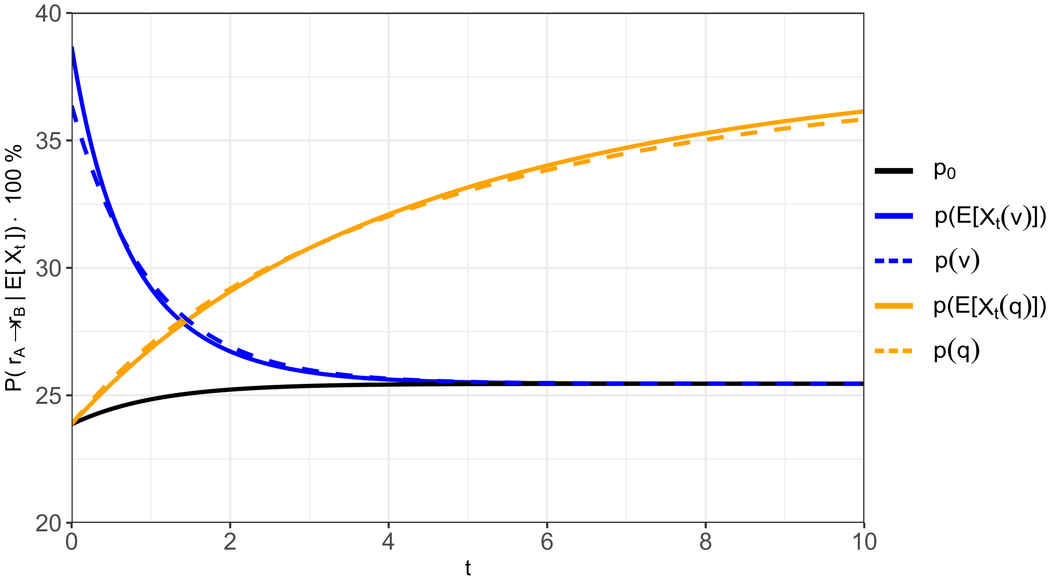


Figure 7: Regime shift probability over time for two types of management actions. Solid lines indicate calculated probabilities, dashed curves are approximations. Parameter values: $v = -25, \Delta p(v) = 0.125, q = 0.2, \Delta p(q) = 0.12, X_0 = 75, \mu_A(c_t) = 90 - 25c_t, \mu_*(c_t) = 80c_t, c_0 = 0.6, \gamma = 0.2, \theta = 1, \sigma = 5, \lambda = 0.3, \bar{y} = -10, \beta = 5$.

Figure 7 shows the fit of the approximation to the actual, calculated probability. The probabilities due to action resemble simple exponential convergence and decay processes because by Equation (12), the expected ecosystem state responds exponentially with rate θ to changes in initial value (action v) and deterministic equilibrium value. The latter is determined by environmental conditions, which change exponentially (action q) with rate γ as given by Equation (17). Due to the number of different parameters that may be affected by actions of type z, we do not provide a general approximation for management type z here.

A different and for some applications more useful way to assess the probability of regime shift is to calculate the probability of one shift within a time interval $[s, t]$ of arbitrary length. This is possible if the value of X_s at time s is known. The relevant time interval for ecosystem management based on probabilistic information is $[0, t]$. In the limit case of $[t, t + \mathrm{d}t]$, this reduces to the instantaneous probability of regime shift given by (25). More generally, the probability can be calculated for any s, t using the transition probability density function $P_t(X)$ of the stochastic process X_t. This density function can be obtained by solving the corresponding Fokker-Planck equation (in shorthand notation)

$$\frac{\partial P_t(X)}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 P_t(X)}{\partial X^2} - \theta\mu\frac{\partial P_t(X)}{\partial X} + \theta\frac{\partial X P_t(X)}{\partial X} - \lambda P_t(X) + \lambda\int_0^\infty P_t(X-y)\,Q(y)\,\mathrm{d}y\,, \quad (28)$$

where $Q(y)$ is the probability distribution function of the jump size y. It is not possible to solve this equation analytically; numerical approximation methods are required to obtain the density function (Gaviraghi, 2017).

2.5 Discussion and conclusions

We have constructed a generic model of ecosystems with alternative stable states and stochastic dynamics, and their management. Our original contribution was to combine a novel deterministic multistability mechanism with two different stochastic influences: continuous diffusion and discrete jumps. Thus, we have improved the representation of stochasticity in models of ecosystems with alternative stable states. This provides a better understanding of the role of different deterministic and stochastic mechanisms and their interaction in causing regime shifts.

We now discuss limitations and potential extensions of the model. First, the model is formulated in terms of a single state variable to establish a clear focus on how stochasticity interacts with deterministic mechanisms of multistability. This neglects potential interactions between multiple state variables which may be relevant for some ecosystems. For some of these systems, it may be possible to construct an index of the ecosystem state (e.g. Blenckner et al., 2021), so that X_t is the index value at time t.

Second, the linearity of X_t in Equation (1) is seemingly at odds with the abrupt and nonlinear nature of regime shifts. The nonlinearity in our model arises from the bistability mechanism in Equation (3) which entails a discontinuous shift in the deterministic equilibrium $\mu(c)$ attracting the state variable. Hence, even though the response of the state variable to changes in its equilibrium value is linear, the overall system dynamics are nonlinear.[9]

Third, we assume that only the location of the deterministic equilibrium $\mu(c)$ changes when a regime shift occurs. We neglect that other parameters (listed in Table 1) could change as well. This is to focus on the core dynamic mechanism of alternative stable states. While it is plausible and easy to integrate in the model that other parameters change, this would not qualitatively change the dynamics of regime shifts.[10]

Last, in our model the uncertainty regarding the dynamics of the ecosystem is probabilistic. That is, we assume perfect knowledge about the distribution of stochastic perturbations and no fundamental uncertainties regarding the location of thresholds, consequences of management actions, or values of model parameters. Essentially, our model is rich in environmental *risk*, but assumes a high degree of knowledge. Depending on the specific system under study, consideration of deeper forms of uncertainty might be needed. This would require a completely different approach to modeling.

With these limitations and reservations in mind, applying the model to ecosystems with alternative stable states as outlined in Section 2.4 opens new pathways for assessing management when stochastic influences are important.

Acknowledgments

We are grateful to Luc Doyen, Elias Ehrlich, Christian Mittelstaedt and two anonymous reviewers for discussion and comments, to Mirko Stappert and Dominik Sulz for providing orientation in the mathematical jungle, and to the German Federal Ministry of Education and Research (BMBF) for financial support under grant no. 01LC1826C. Open Access funding enabled and organized by Projekt DEAL.

[9] In cases where the response of the state variable to changes in its equilibrium value is nonlinear, one can linearize the dynamics around the equilibrium using a first-order Taylor approximation. That is, one may approximate some nonlinear dynamics $F(X_t, c)$ around the equilibrium $X_t = \mu(c)$ so that $F(X_t, c) \approx F_X(\mu(c), c) \cdot (X_t - \mu(c))$.

[10] For instance, if discontinuous jumps represent fire disturbances in a savannah, the jump parameters λ, $\bar{y}$ and β should depend on the vegetation regime to consider fuel available for fires (D'Odorico et al., 2006).

Appendix A Mathematical derivations

A.1 Solution of Equation (6)

Starting with the stochastic differential equation

$$\mathrm{d}X_t = \theta(\mu(c) - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t + y\,\mathrm{d}N_t \tag{A.1}$$

we employ the method of variation of parameters by setting $Y_t = X_t\mathrm{e}^{\theta t}$. Employing Itô's Lemma and the chain rule of differentiation we get:

$$\begin{aligned}\mathrm{d}Y_t &= \theta X_t\mathrm{e}^{\theta t}\,\mathrm{d}t + \mathrm{e}^{\theta t}\,\mathrm{d}X_t\\ &= \theta X_t\mathrm{e}^{\theta t}\,\mathrm{d}t + \mathrm{e}^{\theta t}\left[\theta(\mu(c) - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t + y\,\mathrm{d}N_t\right]\\ &= \mathrm{e}^{\theta t}\theta\mu(c)\,\mathrm{d}t + \mathrm{e}^{\theta t}\sigma\,\mathrm{d}W_t + \mathrm{e}^{\theta t}y\,\mathrm{d}N_t\,.\end{aligned} \tag{A.2}$$

Integrating from 0 to t and using the initial value $Y_{t=0} = Y_0 = X_{t=0}\mathrm{e}^{\theta t} = X_0\mathrm{e}^{\theta t}$, we obtain:

$$Y_t = \left[\mathrm{e}^{\theta t}\mu(c)\right]_0^t + \sigma\int_0^t \mathrm{e}^{\theta s}\,\mathrm{d}W_s + \int_0^t \mathrm{e}^{\theta s}y_s\,\mathrm{d}N_s + K\,. \tag{A.3}$$

Seeing that for $t = 0$, $K = Y_0$, we write:

$$Y_t = Y_0 + \mu(c)\left(\mathrm{e}^{\theta t} - 1\right) + \sigma\int_0^t \mathrm{e}^{\theta s}\,\mathrm{d}W_s + \int_0^t \mathrm{e}^{\theta s}y_s\,\mathrm{d}N_s\,. \tag{A.4}$$

Transforming back with $X_t = Y_t\mathrm{e}^{-\theta t}$, we get the solution in terms of stochastic integrals which is given in the main text:

$$X_t = X_0\mathrm{e}^{-\theta t} + \mu(c)\left(1 - \mathrm{e}^{-\theta t}\right) + \sigma\int_0^t \mathrm{e}^{\theta(t-s)}\,\mathrm{d}W_s + \int_0^t \mathrm{e}^{\theta(t-s)}y_s\,\mathrm{d}N_s\,. \tag{A.5}$$

A.2 Expected value of jumps

The expected value of the last term of (A.5) is obtained as follows:

$$\begin{aligned}\mathbb{E}_{y,\mathrm{d}N}\left[\int_0^t \mathrm{e}^{-\theta(t-s)}y_s\,\mathrm{d}N_s\right] &= \bar{y}\,\mathbb{E}_{\mathrm{d}N}\left[\int_0^t \mathrm{e}^{-\theta(t-s)}\,\mathrm{d}N_s\right]\\ &= \bar{y}\int_0^t \mathrm{e}^{-\theta(t-s)}\lambda\,\mathrm{d}s = \bar{y}\left[\frac{1}{\theta}\mathrm{e}^{-\theta(t-s)}\lambda\right]_0^t\\ \bar{y}\left[\frac{\lambda}{\theta} - \frac{\lambda}{\theta}\mathrm{e}^{-\theta t}\right] &= \bar{y}\,\frac{\lambda}{\theta}\left(1 - \mathrm{e}^{-\theta t}\right)\,.\end{aligned} \tag{A.6}$$

To check whether this expression increases or decreases in θ take the derivative with respect to θ:

$$\frac{\partial}{\partial\,\theta}\bar{y}\,\frac{\lambda}{\theta}\left(1-\mathrm{e}^{-\theta t}\right)=-\bar{y}\,\frac{\lambda}{\theta^2}+\bar{y}\,\frac{\lambda}{\theta^2}\mathrm{e}^{-\theta t}+\bar{y}\,\frac{\lambda}{\theta}t\mathrm{e}^{-\theta t}=-\bar{y}\,\frac{\lambda}{\theta^2}+\bar{y}\,\frac{\lambda\mathrm{e}^{-\theta t}(\theta t+1)}{\theta^2}\,. \tag{A.7}$$

Whether this derivative is positive or negative depends on the sign of $\bar{y}$. For $\bar{y}>0$, the derivative is negative, for $\bar{y}<0$ it is positive. That is, the derivative will be negative if:

$$\bar{y}\,\frac{\lambda}{\theta^2}>\bar{y}\,\frac{\lambda\mathrm{e}^{-\theta t}(\theta t+1)}{\theta^2}\,. \tag{A.8}$$

For $\bar{y}>0$, we have that

$$\begin{aligned}1&>\mathrm{e}^{-\theta t}(\theta t+1)\\ \mathrm{e}^{\theta t}&>\theta t+1\,,\end{aligned} \tag{A.9}$$

which holds by the power series definition of the exponential function for all $t>0$ (since $\theta>0$ by assumption):

$$e^{\theta t}=1+\theta t+\frac{(\theta t)^2}{2!}+\frac{(\theta t)^3}{3!}+\ldots>1+\theta t\,. \tag{A.10}$$

For $\bar{y}<0$, all inequality signs are reversed and the modified form of (A.9) does not hold. The proof for the opposite case of (A.7)<0 is analogous.

Chapter 3

Quantifying agents' responsibility: a generalized measure of causation in dynamical systems

This chapter was written with Stefan Baumgärtner.[*]

This chapter has been submitted for publication in *Ecological Economics* and is currently under review. It has been uploaded as a working paper at: https://ssrn.com/abstract=4277765.

Abstract: How to ascertain causal relationships has been a key question in science and philosophy for centuries. Based on established principles of causation, we develop a quantitative measure of an agent's causal responsibility for the state of a dynamical system: we measure the degree to which an agent's action has caused the system state at a later point in time as the degree to which the action is necessary and sufficient for this state. Our concept can be applied in deterministic as well as in stochastic systems, and for continuous and discrete conceptions of the system state. We find that the extent of causal responsibility crucially depends on the specifics of system dynamics, type of action and the point in time at which the system state occurs. Quantitatively measuring causation in dynamical systems is relevant for attributing an observed system state to its causes, assessing the effectiveness of management actions and policies, or designing liability regulations. Our concept also provides information about the temporal extent of an agent's causal efficacy and, hence, the temporal limits of the agent's normative responsibility.

[*] Author contributions:
Michael Stecher: conceptualization (equal), formal analysis (lead), methodology (lead), software (lead), visualization (lead), writing – original draft (lead), writing – review & editing (equal).
Stefan Baumgärtner: conceptualization (equal), funding acquisition (lead), methodology (supporting), supervision (lead), writing – review & editing (equal)

3.1 Introduction

Many natural and human-made systems are inherently dynamic in the sense that their state and structure change over time. In a dynamical system, the consequences of an agent's action may not become apparent immediately, but only take effect at a later point in time and may be co-determined by natural dynamics. For instance, the discharge of pollutants by a mining company into a river may not have an immediate effect on the river ecosystem, but may – in combination with high water temperatures – facilitate a bloom of toxic algae that leads to a collapse of the fish population in the river weeks after the discharge. To determine who is to blame for the collapse, one needs to know what has caused it. Other than the mining company's discharge, temperature conditions, chance influences, or a combination of these factors could have also played a role in causing the collapse. In such a situation, the challenge is to quantitatively assess to what extent the collapse has been caused by the mining company's discharge – rather than by other factors. This is the mining company's *causal responsibility* for the collapse.

In general, this raises the question of how to measure causation in dynamical systems. More precisely, one would like to know to what extent the system state at a particular point in time can be attributed to an agent's prior action. Further, to evaluate and inform decision-making, one would like to assess an action's effectiveness to reach a given target state as well as its expected causal impact in the future. These questions are relevant in all kinds of dynamical systems that are affected by human actions, including fisheries, forests, agricultural systems, the global climate system, public health, epidemics and vaccination campaigns, financial markets, or the macroeconomy.

In this paper, we develop a measure of an agent's causal responsibility[11] for the state of a dynamical system based on the agent's action and its impact on the subsequent system dynamics. In addition, we study how causal responsibility evolves over time, and how this depends on the type of dynamical system and action.

A number of approaches of how to ascertain and measure the strength of causal relationships exist in the literature. A fundamental distinction between approaches is whether, for a given causal relationship, one aims at identifying the effects of a given cause (e.g., health consequences of a particular lifestyle) or the causes of a given effect (e.g., risk factors for a particular disease) (Holland, 1986). Both perspectives provide valid insights into the

[11] To say that an agent is causally responsible for a system state goes beyond ascertaining that the agent's action has caused the outcome. Agents can only be causally responsible for an outcome if they can choose freely from a range of alternatives that differ qualitatively in their foreseeable consequences (Bovens, 1998). Causal responsibility is purely descriptive and distinct from other layers of responsibility, such as *normative responsibility* – how one should act given some normative framework (Baumgärtner et al., 2018).

causal relationship under study and are relevant for answering different questions. Here, we elaborate the dynamic aspect of the second approach. Before going into the details of this approach, we briefly discuss the main exponent of the first approach, causal inference, as well as concepts that bridge both approaches.

Causal inference in economics and other disciplines measures the effect of a given cause ("treatment") as the difference between two potential outcomes of some response variable: exposure to the treatment versus no exposure ("control") (Haavelmo, 1943; Rubin, 1974; Holland, 1986; Angrist & Pischke, 2009). This basic idea, originally developed for randomized experiments (Neyman, 1923, translated and reprinted in Neyman, 1990), has been extended to identify causal effects using non-randomized empirical data. The "fundamental problem of causal inference" (Holland, 1986) that both treatment and control cannot be observed on the same unit is overcome by considering the average treatment effect over a larger population of units. In dynamic settings, the time-varying causal effect of a factor can be measured as the cumulative average treatment effect over time (Jordà et al., 2022). The validity of causal inferences rests on several assumptions about the data-generating process and the suitability of the chosen identification strategy, including the "stable unit treatment value assumption" (SUTVA)[12] and "excludability"[13]. In coupled human and natural systems violations of SUTVA and excludability are likely, which may bias causal inferences (Ferraro et al., 2019).

A more basic perspective on causal relationships in dynamic settings that bridges the two approaches – effects of a given cause or causes of a given effect – analyzes whether two factors are causally related at all. This approach, of which the most prominent exponent is known as "Granger causality" (Granger, 1969), is based on the notion of predictability: one time-series variable is said to "Granger cause" a second one if it improves the ability to forecast future values of the second. Hence, Granger causality reflects whether two variables are "temporally related" (Granger & Newbold, 1977), but provides no information on the strength or nature of the underlying causal relationship. In ecosystems and other nonseparable weakly coupled dynamical systems, where Granger causality is not applicable, a similar approach was suggested by Sugihara et al. (2012). Their methodology based on convergent cross mapping is useful to identify whether two species in an ecosystem do or do not interact, but cannot be used to attribute a particular ecosystem state to various factors, including agents' actions.

[12] SUTVA states that there is no interference between units in the sense that the outcome of treatment in one unit depends on the treatment of other units (Rubin, 1980).

[13] Excludability states that unobserved heterogeneity arising from confounding factors that drive variation in the response variable beyond their effect on treatment has been accounted for by an adequate treatment assignment mechanism (Ferraro et al., 2019).

There is a rich and long-standing literature that aims at identifying the causes of a given effect (e.g. Hume, 1739; Mill, 1843; Wright, 1921; Reichenbach, 1956; Bunge, 1959; Hart & Honoré, 1959; Good, 1961; Mackie, 1965; Lewis, 1973; Pearl, 2009b). This literature has largely focused on the conditions under which an action is considered a cause of an outcome, and when it is not. That is, causation is typically understood in a binary sense rather than as a cardinal measure of the degree to which a given outcome was caused by one cause relative to another. There are a number of contributions developing such a cardinal measure.

Vallentyne (2008) proposes a measure of an agent's "partial responsibility" for an outcome based on the increase in the outcome's probability that is directly and indirectly due to the agent's action. This achieves a full attribution of causality, but only considers a single outcome in a highly stylized probabilistic system. Pearl (2009b) proposes separate measures for the "probability of necessity", "probability of sufficiency" and "probability of necessity and sufficiency" relating two binary variables. This is based on the distinction (Mackie, 1965; Mitroff & Silvers, 2013) between necessary causation (i.e., the outcome could not have occurred without the cause) and sufficient causation (i.e., the cause was, all by itself, capable of producing the outcome). Which of these is an adequate measure of causation may depend on the context (Hannart et al., 2016). While explicating the concepts of necessary and sufficient causation in a probabilistic context, Pearl's (2009b) approach does not ascribe causality to agents and their actions. Gleiss & Schemper's (2019) measures for a prognostic factor's "degree of necessity" and "degree of sufficiency" in an epidemiological context are similar and do not refer to agency either.

Empirical work on the degree of causation has recently gained attention in the context of extreme event attribution in climate science (Allen, 2003; Stott et al., 2004; Otto, 2017). There, the question is to what extent a particular climatic event can be attributed to anthropogenic greenhouse gas emissions rather to natural climate variability. The answer to this question is given by the relative increase in the likelihood of the event compared to a counterfactual climate without anthropogenic forcing, which essentially measures how necessary climate change is for the occurrence of this event. The event to be attributed needs to be defined in terms of a threshold of a climatic variable (e.g., a heatwave is defined as the monthly average temperature in a particular region exceeding a certain value), which may be "to a large extent arbitrary" (Hannart et al., 2016).

Questions of causal attribution have also been discussed in the context of material flow analysis, for instance, how to measure the responsibility of consumers and producers for greenhouse gas emissions caused by the production and consumption of goods (e.g., Bastianoni et al., 2004; Rodrigues et al., 2006; Lenzen et al., 2007). While it is a strength of this literature that material flows are attributed to different agents, these treatments are deficient in several ways. First, the proposed measures are largely ad hoc and not systematically based on principles of causation. Second, the notion of “responsibility” employed in this literature confounds descriptive aspects of causation and normative aspects of fairness

A handful of contributions are concerned with determining the relative causal contributions of individual agents in situations where an outcome is jointly caused by the simultaneous actions of multiple agents. Chockler & Halpern (2004) propose a measure based on contingency, which captures how many changes need to be made to the circumstances before an action makes a critical difference for the outcome. Their concept of “degree of responsibility” can lead to considerable over- or underattribution of causality. Braham & van Hees (2009) measure an action’s degree of causation as the relative frequency in which the action is a necessary element of a set of conditions which is jointly sufficient for the outcome. This avoids over- or underattribution, but is not applicable in a stochastic system where the outcome consists of infinitely many potential realizations of the continuous system state. Mittelstaedt & Baumgärtner (2023) measure an agent’s individual causal responsibility as the marginal increase in the outcome’s probability due to the agent’s action averaged over all hypothetical sequences in which the simultaneous actions of all agents might unfold. This achieves a full attribution of causality in a stochastic system, but is limited to dichotomous outcomes in systems with two discrete states.

Our novel contribution here is to develop a generalized measure of the degree of causation of a given outcome by an agent’s action in a dynamical system. Specifically, we measure an agent’s causal responsibility for the realized state of a dynamical system as the degree to which the agent’s action is necessary and sufficient for this state. Our concept is founded upon established principles of causation and achieves a full attribution of causality that is consistent across deterministic and stochastic systems for both discrete and continuous conceptions of the system state. Furthermore, we study how the agent’s causal responsibility evolves over time for different types of actions and systems. This is relevant for a number of applications in which an action’s consequences dynamically unfold in a non-trivial way. For instance, our concept can be used for attributing a realized system state to its causes, assessing the effectiveness of management actions for given goals, designing economically efficient liability regulations, and quantifying the temporal limits of normative obligations.

This paper is organized as follows. In Section 3.2 we present a simple and general setup of stochastic dynamical systems, which forms the basis of our analysis. In Section 3.3 we review established philosophical ideas on causation and develop a quantitative measure of causal responsibility. In Section 3.4 we apply this measure to a number of dynamical systems and different management actions. In Section 3.5 we highlight the relevance of our concept and its implications for normative responsibility. In the final Section 3.6 we discuss limitations and conclude.

3.2 Model and setup

The evolution of the system state[14] $X_t \in [0, \infty)$ over time $t \in [0, \infty)$ is described by a stochastic differential equation of form

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}Z_t \ , \tag{29}$$

where $f(\cdot)$ and $g(\cdot)$ are continuously differentiable functions and Z_t is some stochastic process. The known initial value of X_t at $t = 0$ is x_0. In deterministic systems, $g(X_t) = 0$ for all X_t. The state of the system at any point in time can be obtained by solving Equation (29) analytically or numerically. Suppose that the solution over the entire time interval $[0, \infty)$ is known. We assume that the stochastic process X_t (Equation 29) satisfies the Markov property and converges to a stationary probability distribution in finite time.

Given the stochastic dynamics (29) of the system state X_t, there exists an unconditional probability density function $\rho_{X_t}(x)$. Hence, the probability that X_t lies in the interval $[\underline{x}, \bar{x}] \subseteq [0, \infty)$ is given by:

$$P\big(X_t \in [\underline{x}, \bar{x}]\big) = \int_{\underline{x}}^{\bar{x}} \rho_{X_t}(x)\,\mathrm{d}x \ . \tag{30}$$

Conditioned on the initial value x_0 at time $t = 0$, there exists a conditional probability density function $\rho_{X_t|x_0}(x)$. The conditional probability that X_t lies in the interval $[\underline{x}, \bar{x}] \subseteq [0, \infty)$ given the initial value x_0 is thus:

$$P\big(X_t \in [\underline{x}, \bar{x}] \ \big| \ x_0\big) = \int_{\underline{x}}^{\bar{x}} \rho_{X_t|x_0}(x)\,\mathrm{d}x := p(X_t, \underline{x}, \bar{x}) \ , \tag{31}$$

where the last expression is introduced to simplify notation.

Actions

There is a single agent that takes a one-time action a at time $t = 0$ which modifies the dynamics of X_t:

$$\mathrm{d}X_t^a = f(X_t, a)\,\mathrm{d}t + g(X_t, a)\,\mathrm{d}Z_t \ . \tag{32}$$

Consequently, the probabilities (30) and (31) are also modified. We assume that the agent knows these probabilistic consequences of acting.

In principle, an action could modify the initial system state x_0, the deterministic drift

[14] For systems with multiple state variables, it may be possible to construct an index of the ecosystem state, so that X_t is the index value at time t.

f of the process or its stochastic factor g. Specifically, we consider the following distinct types of management actions that affect the probability distribution of X_t in different ways. These action types are idealized cases that, in reality, may occur in combination or come in different variants. We restrict our analysis to actions that change the moments of the distribution of the process, but not its existence or stationarity.

i) **Initial value modification:** $x_0 \neq x_0^a$

Modifying the initial value of the process directly and instantaneously changes the system state. This changes the conditional probability density $\rho_{X|x_0^a}(x)$. Examples include extracting a certain amount of a natural resource (e.g., clear-cut harvesting of timber) or replenishing its stock (e.g., afforestation).

ii) **Drift modification:** $\mathrm{d}X_t^a = f(X_t, a)\,\mathrm{d}t + g(X_t)\,\mathrm{d}Z_t$

Modifying the deterministic drift may affect the probability distribution in two different ways: we distinguish between attractor modifications, which change the mean $\mathbb{E}[X_t]$ of the unconditional distribution, and rate modifications, which do not.

a) **Attractor modification:** $\mathbb{E}[X_t^a] \neq \mathbb{E}[X_t]$

Modifying an attractor changes the mean of the unconditional distribution of X_t^a (i.e., the value X_t^a converges against in the long run). In ecological systems, this corresponds to modifying the carrying capacity of a population, for instance by changing resource availability or trophic interactions (e.g., removing competitors or introducing alien species).

b) **Rate modification:** $\mathbb{E}[X_t^a \,|\, x_0^a] \neq \mathbb{E}[X_t|x_0], \quad \mathbb{E}[X_t^a] = \mathbb{E}[X_t]$

Rate modifications change the conditional mean, but do not affect its unconditional mean. In particular, rate modifications alter the speed and variability of the convergence process towards the unconditional distribution. In ecological systems, this affects the return time to equilibrium after a perturbation, which is known as stability (Holling, 1973) or engineering resilience (Pimm, 1984). In technical and biochemical systems, this corresponds to catalyzing a reaction or accelerating bacterial growth through higher ambient temperature.

iii) **Volatility modification:** $\mathrm{d}X_t^a = f(X_t)\,\mathrm{d}t + g(X_t, a)\,\mathrm{d}Z_t$

Modifying the stochastic factor g of the process changes the susceptibility of the system state to stochastic influences. This primarily changes the variance and higher moments of the conditional and the stationary distribution of the process. In agricultural systems, constructing irrigation infrastructure or dams insures the crop output against adverse environmental fluctuations such as drought or flooding.

iv) **Choice of control strategy:** $\mathrm{d}X_t^a = [f(X_t) - a(X_t)]\,\mathrm{d}t + g(X_t)\,\mathrm{d}Z_t$

Choosing a particular control strategy at time $t = 0$ continuously, at each time t, reduces or increases the stock by a certain amount $a(X_t)$. Examples include continuous harvesting of a renewable natural resource (e.g, exploiting a fish stock) or the emission of pollutants (e.g. greenhouse gases or nutrients from fertilizer use). This changes mean and higher moments of both the conditional and the unconditional distribution of the process. We consider three different types of control strategy:

a) **Constant amount:** $a(X_t) = h$
Extracting a constant amount h at each time t, irrespective of the stock level, can be thought of, e.g. as harvesting for subsistence.

b) **Constant fraction:** $a(X_t) = hX_t$
Extracting a constant fraction h of the current stock level, i.e. extracting more when the stock level is high and less when it is low, can be thought of as a rudimentary adaptive harvesting strategy.

c) **Intertemporally optimal amount:** $a(X_t) = h^*(X_t)$
Extracting, at each time, the amount $h^*(X_t)$ that solves some biological or economic optimization problem, e.g. maximization of welfare or net benefits subject to ecological constraints.

We study the effects of these idealized action types with illustrative and practically relevant examples in Section 3.4.

3.3 Conceptualizing and measuring causal responsibility

Causal responsibility ascribes the consequences of an action to its perpetrator.[15] In a dynamical system, the consequences of an action consist of subsequent system states which result from the modified system dynamics due to action. Which consequences are to be ascribed to the actor needs to be specified and may be conceptualized in different ways. In principle, one ascribes the realized system state at a particular point in time being in a specific interval, where the interval and the point in time are to be specified ("causal responsibility

[15] We use the term "causal responsibility" here for what is also known as "ascriptive responsibility" (Baumgärtner et al., 2018) or "agent-responsibility" (Vallentyne, 2008).

for what?"). Causal responsibility is purely descriptive and independent of any norm about how the system state ought to be or what action ought to be taken.

A quantitative measure of causal responsibility should satisfy a number of principles of causal attribution. In the the next subsection, we discuss these principles and what they imply for the quantitative measurement of causal responsibility. Subsequently, we suggest a measure that fulfills these principles. First, we present the simplified version for deterministic systems before presenting the generalized measure for stochastic systems.

3.3.1 Principles of causal attribution

To substantiate the meaning of causal responsibility, we start from general and accepted ideas on causation. In particular, we discuss:

1. counterfactual causation

2. necessary and sufficient causation

3. multiple causes

4. singular vs. general causation (ex-post vs. ex-ante perspective)

While these are not independent, we discuss them in turn. To start with, we employ an ex-post perspective, meaning that we start from the singular case of an actually realized system state and retrospectively ask about its causes. We explicitly consider the aspect of taking an ex-ante vs. an ex-post perspective when discussing point 4.

Counterfactual causation

We employ a counterfactual conception of causation that may be summarized as: "we think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it" (Lewis, 1973, p. 557). Clearly, an action did not cause a particular system state if the action did not make a difference for this system state to occur compared to the counterfactual of not acting. Using a counterfactual approach is only possible in a system in which causal relationships can be identified and described through a predictive model (Pearl, 2009b, Chap. 7), such as in Section 3.2. This conception of causation implies three important properties of causal responsibility:

i) An agent's causal responsibility is measured relative to the reference scenario of not acting.

ii) An agent's causal responsibility for the system state at time t is different for two different actions taken under the same circumstances (and hence the same counterfactual system state) if and only if the actions entail (probabilistically) different system states at time t. And the larger the difference in the (probability of the) resulting system states, the larger the difference in causal responsibility.

iii) An agent's causal responsibility for the system state at time t when taking a given action may be different under different circumstances. That is, an agent's causal responsibility does not only depend on the action taken, but also on the circumstances under which the action's consequences unfold (and which also modify the counterfactual system state).

Necessary and sufficient causation

In general, one distinguishes between necessary and sufficient causation (Mackie, 1965; Braham & van Hees, 2009; Pearl, 2009b; Mitroff & Silvers, 2013; Gleiss & Schemper, 2019). A cause is *necessary* for an outcome if the outcome would not have occurred in the absence of the cause. This notion of "but for" causation is predominant in the law (Hart & Honoré, 1959; Hannart et al., 2016) and captures one important condition of causation, but does not by itself imply that the outcome actually occurs. The other important aspect is sufficiency: a cause is *sufficient* for an outcome if the outcome must occur in the presence of the cause. An outcome is fully determined by a cause if and only if the cause is both necessary and sufficient for the outcome. Hence, the attribution of causal responsibility for an outcome to an agent should be based on necessary and sufficient causation.

Multiple causes

Typically, there are multiple causes for an outcome. In our setting (Section 3.2), a given system state may be caused by the agent's action, or natural dynamics, or a combination of both. Hence, an action may not be entirely necessary and sufficient for a given system state, but only partially (Chockler & Halpern, 2004; Vallentyne, 2008; Braham & van Hees, 2009). Thus, an agent's causal responsibility for the realized system state should measure the *degree* to which the agent's action is necessary and sufficient for this state. Likewise, the degree to which natural dynamics are necessary and sufficient for the realized system state is attributed to "nature". The agent's causal responsibility and the causality attributed to nature should add up to one, so that the actual system state is fully and disjointly explained by its causes. This guarantees that there is neither over- nor underattribution.[16]

[16] Overattribution means that the sum of causal responsibility attributed to individual causes is greater than 1. It typically arises from causal overdetermination, which occurs when multiple causes are present,

Regarding sufficient causation with multiple causes, natural dynamics are completely sufficient for the counterfactual system state. In turn, the agent's action is completely sufficient for the difference between the realized and the counterfactual system state. Hence, both are only partially sufficient for the realized system state at a particular point in time: the degree to which natural dynamics are sufficient is given by the relative contribution of the counterfactual to the realized system state; the action's degree of sufficiency is given by the relative difference in state that the action makes.

In stochastic systems, natural dynamics also include random fluctuations of the system state. In our setting, this implies that for a given action any system state may occur with some probability. Hence, no action can be completely necessary for a realized system state, because there is always the possibility that this system state is realized by pure chance in the absence of action. The degree to which an action is necessary for a realized system state is given by the change in the state's probability due to action compared to the counterfactual probability entailed by not acting. The larger the increase in probability due to action, the larger is the action's degree of necessity. The action is completely unnecessary for a realized system state if it does not increase, or if it decreases, the state's probability of occurring.

Singular vs. general causation (ex-post vs. ex-ante perspective)

So far, we have considered a realized system state at some point in time and retrospectively asked about its causes. This is an ex-post perspective, which is adequate for a particular outcome that has actually occurred ("singular causation") (Pearl, 2009b). Alternatively, one may ask prospectively[17] at the time of action about the action's *expected* causal impact on the future system state. This is an ex-ante perspective, which is adequate for the general tendency of an action to bring about some outcome that might occur in the future ("general causation") (Mackie, 1965).

In deterministic systems, both perspectives are equivalent. In stochastic systems, which perspective is used when attributing causality makes a conceptual difference. When taking an ex-post perspective, one only considers a single random realization of the system state – and none of the infinitely many other potential states that could have been realized at that time. When taking an ex-ante perspective, forming an expectation about the action's consequences requires that one considers all potential system states at that time.

of which any one would be entirely sufficient for the outcome individually, such as when a victim dies from being shot simultaneously by multiple assassins. In criminal law, overattribution may be desirable – *all* the assassins are legally fully responsible for the victim's death (cf. Hart & Honoré, 1959; Honoré, 1995).

[17] Both the retrospective and prospective assessment discussed here are purely descriptive. In particular, the prospective assessment is not normative (what one *should* do), and the retrospective is not judging (how one should have acted) (cf. Baumgärtner et al., 2018, Sec. 3.2).

Against this background, an agent's ex-post causal responsibility is an answer to the question: "To what extent has the agent's action a at time 0 caused the realized system state at time t?" In contrast, an agent's ex-ante causal responsibility answers a different question, namely: "To what extent can the agent's action a at time 0 be expected to cause the resulting system state at time t?" The ex-ante causal responsibility is simply the expected value of the ex-post measure. Both concepts carry different information about an action's causal efficacy and are relevant for different purposes. The ex-post concept is the relevant one to attribute a specific realized system state to its causes ("singular causation"). It is thus subject to the randomness inherent in the realization of a particular system state. The ex-ante concept, by considering all potential realizations, reveals an action's causal efficacy on the system in a representative manner ("general causation"). While providing general insight on the causal efficacy of an action, it may differ substantially from the ex-post causal responsibility for a particular, random realization.

3.3.2 Causal responsibility in deterministic systems

Suppose the state of some deterministic dynamical system is $x^a_{t_1}$ at time $t_1 \geq 0$ and the agent modified the system dynamics by taking action a at time $t=0$. In this certain environment, both the action and natural dynamics were completely necessary for the realized system state, meaning that $x^a_{t_1}$ could not have resulted at time t_1 without either of them. That is, both the action's degree of necessity and that of natural dynamics are 100%. In line with Section 3.3.1, an agent's causal responsibility measures the degree to which the agent's action is necessary and sufficient for the realized system state. We take the degree of necessary and sufficient causation as the product of the degree of necessity and the degree of sufficiency. Hence, measuring causal responsibility for the state of deterministic systems reduces to measuring an action's degree of sufficiency for the realized system state.

For known deterministic dynamics, an action's degree of sufficiency (and thus an agent's causal responsibility) for the system state $x^a_{t_1}$ at time t_1 is given by the relative difference between the realized and the counterfactual system state x_{t_1} at time t_1. The counterfactual system state that would have resulted in the absence of action a (Figure 8) is uniquely determined by Equation (29) with $g(X_t) = 0$ for all X_t.

Definition 2. An agent's causal responsibility for the realized deterministic system state x^a_t at time t, given the counterfactual system state x_t, and given that the agent took action a at time $t=0$, is given by:

$$R(x^a_t, x_t) = \frac{|\, x^a_t - x_t \,|}{\max\{x^a_t, x_t\}} \tag{33}$$

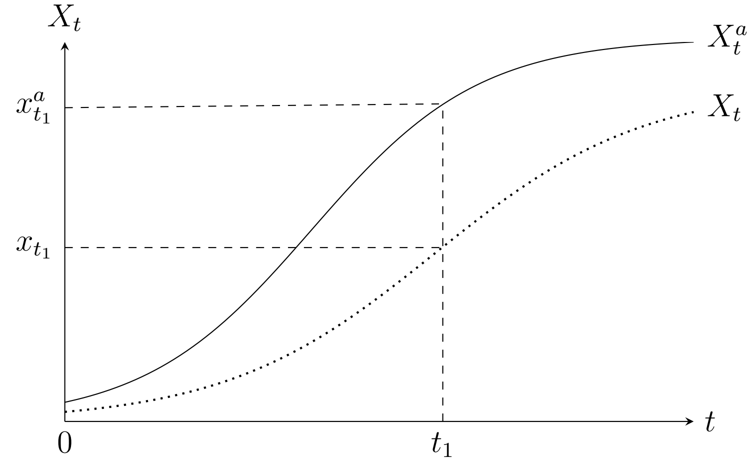


Figure 8: Intuition of measuring causal responsibility in deterministic systems. Actual system state X_t^a (Equation 32 with $g(X_t, a) = 0$) and counterfactual system state X_t (Equation 29 with $g(X_t) = 0$) over time.

The numerator of (33) takes the absolute value of the difference between the realized and the counterfactual system state because it does not matter for causation whether action a increases or decreases the system state relative to the counterfactual. In contrast, the normalization factor in the denominator depends on whether a increases or decreases the system state relative to the counterfactual. It consists of whichever of the two – realized or counterfactual system state – is greater at time t to consistently measure the relative difference to the counterfactual that is due to action. The agent is not causally responsible for the realized system state if the action is completely insufficient for this state, that is, if it does not change the system state relative to the counterfactual. The agent is fully causally responsible if and only if the action is completely sufficient for the resulting system state, which implies that either $x_t^a = 0$ or $x_t = 0$. Between these extreme cases, an agent's causal responsibility lies between 0 and 100%.

The causal responsibility measure (33) has been introduced from an ex-post perspective, but formally also holds for the ex-ante perspective.

3.3.3 Causal responsibility in stochastic systems

In stochastic systems, causality needs to be attributed under uncertainty. One only observes a single random realization X_t^a of the stochastic process X_t^a and none of its infinitely many other potential realizations (Figure 9).[18] In addition, the counterfactual process X_t in the absence of action has also infinitely many other potential realizations.

[18] In slight abuse of notation, we denote the process and its realization by the same variable X_t. Which of the two is meant should be obvious from the context.

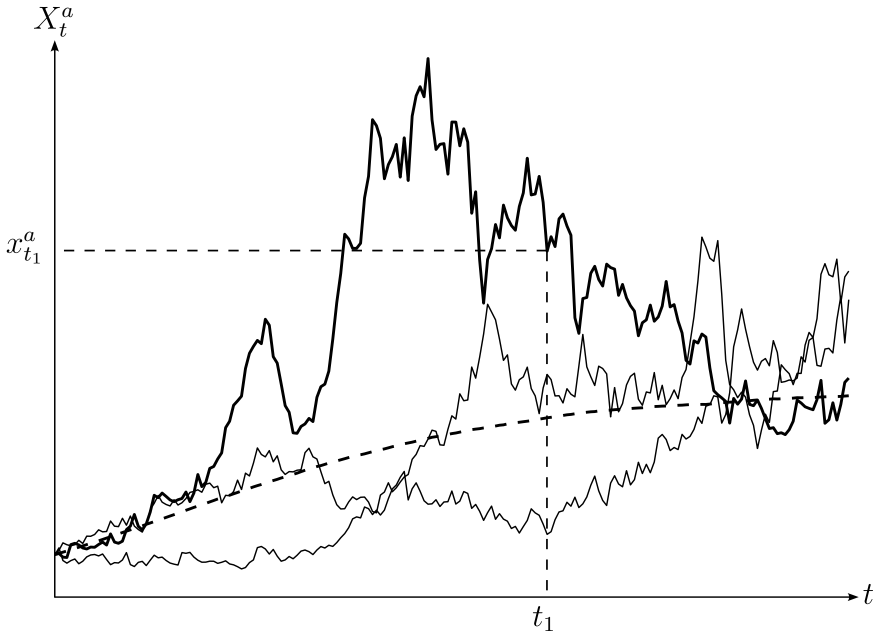


Figure 9: Epistemological problem in stochastic dynamical systems. three random realizations of the stochastic process X_t^a (described by Equation 32) and its expected value (dashed curve). In practice, one only observes a single realization, such as the one drawn in bold, with corresponding system state $x_{t_1}^a$ at time t_1.

Similar to the deterministic case, an action's degree of sufficiency is measured as the relative difference between the realized system state x_t^a and the counterfactual system state x_t. In a stochastic system with known dynamics (32), the counterfactual system state that would have been realized in the absence of action a can be uniquely determined as follows:[19] First, for a given realization X_t^a, the stochastic forcing Z_t apparent in the time evolution of X_t^a is separated from the known deterministic trajectory of the system, by calculating the realization of the stochastic process Z_t from the other known quantities in Equation (32). This particular realization Z_t of the stochastic forcing is then used to simulate the counterfactual realization X_t by inserting Z_t into Equation (29).[20]

Beyond sufficiency, in stochastic systems one also needs to consider how necessary the action was for the realized system state and to what extent there were other potential causes. Measuring an action's degree of necessity requires calculating the probability of finding the process in an interval $[\underline{x}, \bar{x}]$ around the realized system state X_t^a, where $\underline{x}$ and $\bar{x}$ need to be specified. Specifically, we take an action's degree of necessity as the relative difference between two probabilities: the probability $p(X_t^a, \underline{x}, \bar{x})$ of observing the realized system state

[19] We thank Hermann Held for suggesting this procedure to us.

[20] If the stochastic forcing cannot be separated from the deterministic trajectory of the system, but arises, for example, from the high dimensionality of the system dynamics, one needs to use an alternative quantity, such as the expected value of the counterfactual system state in the absence of action.

given the modified process due to action X_t^a, and the probability $p(X_t, \underline{x}, \bar{x})$ of observing this state given the counterfactual process in the absence of action X_t. This measures by how much, in relative terms, the action makes the realized system state more likely.

For illustration, consider two actions a and a' that both increase the probability of the realized system state by the same absolute amount of 30 percentage points, but relative to different counterfactual probabilities in the absence of action $p(X_t, \underline{x}, \bar{x}) = 0.1$ and $p(X_t', \underline{x}, \bar{x}) = 0.6$. The degree of necessity of action a is $(0.4 - 0.1)/0.4 = 0.75$, whereas that of action a' is $(0.9 - 0.6)/0.9 = 0.33$. The former is larger than the latter because the realized system state is made *relatively* more likely by action a – although it is more likely in absolute terms for action a'.[21]

In conclusion, causal responsibility for the state of stochastic systems is determined by the product of two factors: the relative difference between the realized and the counterfactual system state in the absence of action (the action's degree of sufficiency) and the relative difference in the probability of the realized system state (the action's degree of necessity). In line with Section 3.3.1, the degree of necessity, and hence causal responsibility, is zero for any action that does not increase, or decreases, the probability of x_t^a.

Definition 3. An agent's ex-post causal responsibility for the actually realized system state x_t^a at time t, given the probabilistic knowledge available at time $t = 0$, is given by:

$$R\big(x_t^a, x_t\big) = \begin{cases} \dfrac{p(X_t^a, \underline{x}, \bar{x}) - p(X_t, \underline{x}, \bar{x})}{p(X_t^a, \underline{x}, \bar{x})} \cdot \dfrac{\mid x_t^a - x_t \mid}{\max\{x_t^a, x_t\}} & \text{for } \ p(X_t^a, \underline{x}, \bar{x}) > p(X_t, \underline{x}, \bar{x}) \\ 0 & \text{for } \ p(X_t^a, \underline{x}, \bar{x}) \leq p(X_t, \underline{x}, \bar{x}) \end{cases} . \qquad (34)$$

The first factor can also be interpreted in a different manner, namely as a prefactor that measures which part of the relative difference between the realized and the counterfactual system state in the absence of action is attributable to action.[22] The relative difference between the probability due to action and the probability when not acting decreases as the uncertainty surrounding the system dynamics (σ in our setting) increases. That is, the larger the uncertainty, the lower is an action's degree of necessity and thus also causal responsibility. In the extreme case of absolute certainty (i.e., $p(X_t^a, \underline{x}, \bar{x}) = 1$ and $p(X_t, \underline{x}, \bar{x}) = 0$) the causal responsibility measure (34) reduces to the deterministic measure (33) presented in

[21] This is equivalent to the systematic attribution procedure presented by Baumgärtner (2020). In this procedure, a fraction of $\big[p(X_t^a, \underline{x}, \bar{x}) - p(X_t, \underline{x}, \bar{x})/p(X_t^a, \underline{x}, \bar{x})$ of the "outcome luck" (Vallentyne, 2008), i.e., the remaining probability difference $1 - p(X_t^a, \underline{x}, \bar{x})$, is attributed to the agent in addition to the direct probability shift of $p(X_t^a, \underline{x}, \bar{x}) - p(X_t, \underline{x}, \bar{x})$. In the discrete setting studied by Baumgärtner (2020), this is equal to causal responsibility.

[22] In climate attribution science, this factor is known as the "fraction of attributable risk" (Allen, 2003; Jaeger et al., 2008; Otto, 2017; Pfrommer et al., 2019).

Section 3.3.2.

An agent's ex-ante expected causal responsibility is given by the expected value of her ex-post causal responsibility across all potential realized system states weighted by their respective probability of occurring $p(X_t^a, \underline{x}, \bar{x})$. That is, the expected value is calculated with respect to the conditional distribution of the process modified by the action, which represents the agent's state of probabilistic knowledge at the time of action.

Definition 4. An agent's ex-ante expected causal responsibility at time t for taking action a is given by:

$$R^e(a) = \mathbb{E}\big[R(x_t^a, x_t)\big] \ . \tag{35}$$

While the ex-ante expected responsibility is clearly defined, it may not exist in closed-form, but rather has to be obtained through simulations.

3.4 Application and results

In this section, we apply the measures (33), (34) and (35) of causal responsibility to four stylized examples of different dynamical systems covering both deterministic and stochastic dynamics with and without thresholds. These examples have emerged from a more encompassing analysis that we have performed and were chosen because they are well-suited to illustrate the essential results. In Section 3.4.3, we present general results and conjectures that follow from the examples presented here and are also informed by our more encompassing analysis.[23]

3.4.1 Deterministic logistic growth

Consider some renewable resource, such as a fish stock or a forest stand, for which the evolution of the resource stock over time is given by the logistic equation:

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = rX_t\left(1 - \frac{X_t}{K}\right) \ , \tag{36}$$

where the rate of increase of the stock is determined by the intrinsic growth rate r, its maximum stable population size is determined by the carrying capacity K and the initial value is x_0. Equation (36) has a single, stable non-trivial equilibrium at $X_t = K$. In this model, the elementary action types presented in Section 3.2 correspond to modifying, at time 0, the values of r (rate modification) and K (attractor modification), which affect the

[23] For the systems presented here, we studied a range of parameter values and action combinations. We have also studied other types of systems, including the Solow (1956) model of capital accumulation, the Lotka-Volterra model of predator-prey population dynamics (Lotka, 1925; Volterra, 1926), and a model of stochastic ecosystems with alternative stable states (Stecher & Baumgärtner, 2022b).

stock size indirectly, as well as directly modifying the initial value x_0. Control strategies are represented by adding the control term $a(X_t)$ to the right-hand side of Equation (36).

For actions that modify the initial value x_0, such as a one-time reduction or replenishment of the stock of a natural resource, an agent's causal responsibility (Equation 33) for the system state is maximal at time 0 and subsequently decreases over time (Figure 10a). Both the actual system state and the counterfactual system state converge to the same attractor K, only from different initial values x_0 and x_0^a. Hence, causal responsibility for the system state converges to zero over time as the relative difference of the actual to the counterfactual system state in the absence of action decreases to zero. That is, the action's degree of sufficiency decreases over time, whereas natural dynamics become increasingly sufficient for the system state.

For actions that modify the carrying capacity K, for instance by changing resource availability or trophic interactions, an agent's causal responsibility for the system state is zero initially and subsequently increases over time (Figure 10b). As the system state converges to its modified carrying capacity K^a, causal responsibility converges against its maximum level over time. As the actual and the counterfactual system state converge to different attractors from the same initial value, the relative difference between x_t^a and x_t increases over time.

For actions that modify the intrinsic growth rate r, an agent's causal responsibility for the system state is zero at first, followed by a temporary increase, before it subsequently decreases to zero (Figure 10c). Examples include improving the spawning habitat of a fish stock or planting a faster-growing tree species. Both the absolute and the relative difference between the factual and the counterfactual system state increase at first due to the growth rate differential. As both X_t^a and X_t converge to the same attractor K over time, the action's degree of sufficiency subsequently decreases and converges to zero.

For harvesting a constant amount h of the stock at each time, an agent's causal responsibility for the system state is zero initially and subsequently increases over time (Figure 10e). In the depicted case, harvesting follows the maximum sustainable yield (MSY) paradigm, which keeps the stock level constant at its most productive level of half the carrying capacity. Still, causal responsibility increases over time as the counterfactual system state increases over time. In the extreme case of choosing a high harvesting amount that reduces the stock to zero at some point, the agent is fully responsible for the stock collapse from this point on.

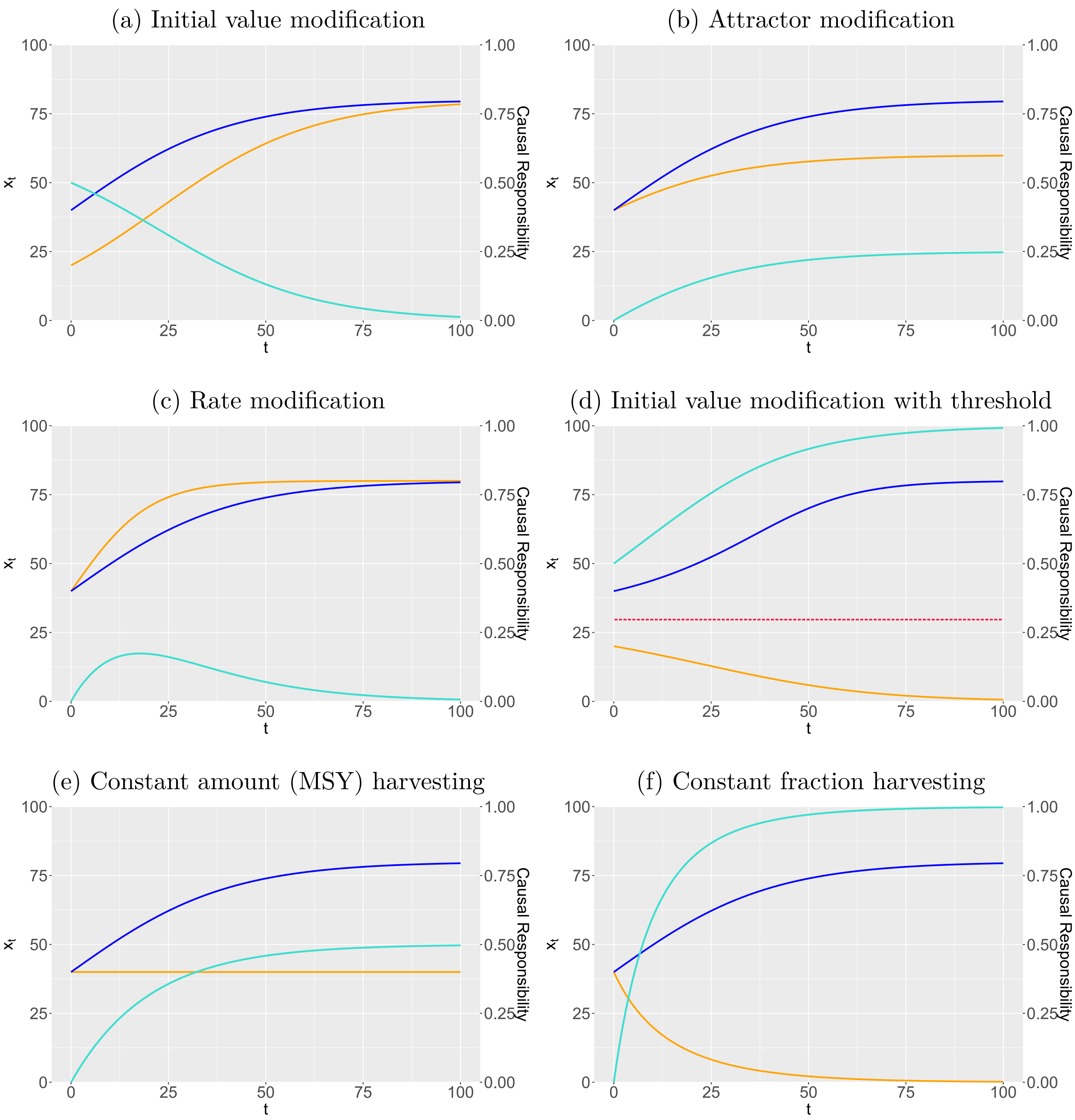


Figure 10: Causal responsibility for different action types under deterministic logistic stock dynamics with and without thresholds. Evolution of the actual system state X_t^a (solid orange), counterfactual system state X_t (solid blue) and causal responsibility $R(x_t^a, x_t)$ (Equation 33) (solid turquoise) over time under deterministic logistic stock dynamics with and without threshold (Equation 36 for a,b,c; Equation 37 for d) for different action types (a-d). Parameter values: $r = 0.05, K = 80, x_0 = 40$ in a-f, $x_0^a = 20$ in a and d, $K^a = 60$ in b, $r^a = 0.1$ in c, $V = 15$ in d (dashed red), $h = 1$ in e, $h = 0.1$ in f.

Similarly, for harvesting a constant fraction h of the stock at each time, an agent's causal responsibility increases and converges to its maximum level over time (Figure 10f). In the depicted case, the agent is eventually fully responsible for completely exhausting the stock by means of choosing an unsustainably high harvesting rate. Conversely, an agent is only partially responsible, i.e. $R(x_t^a, x_t) < 1$, for any system state with a positive stock level, e.g. before the stock is completely exhausted or when choosing a lower harvesting rate that does not exhaust the stock.

Consider now a renewable resource that exhibits critical depensation. That is, the resource stock decreases and converges to zero for stock levels below a critical threshold $V < K$. In ecological systems, this phenomenon of population density being positively related to individual fitness is known as the Allee effect (Allee et al., 1949). Examples include a minimum viable population size necessary for successful reproduction or a minimum level of forest cover that is required for maintaining a suitable microclimate. The dynamics of a resource stock with critical depensation can be described by (Clark, 1990):

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = rX_t\left(1 - \frac{X_t}{K}\right)\left(\frac{X_t}{V} - 1\right) \ . \tag{37}$$

The stability properties of Equation (37) are different from those of Equation (36): in addition to the stable equilibrium at $X_t = K$, Equation (37) has an unstable equilibrium at $X_t = V$. With that, the same actions may entail completely different consequences than without critical depensation.

For actions that reduce the system state below the critical threshold, an agent's causal responsibility for the system state increases and converges to its maximum value of 1 over time. That is, if the threshold is crossed due to the action, the agent is fully responsible for the resulting resource depletion as the action becomes completely sufficient for the system state. This is only possible for actions that directly affect the system state, i.e. initial value modifications or choosing a control strategy. For instance, for reducing the initial value below the critical threshold ($x_0^a < V < x_0$), the agent is fully responsible for the eventual exhaustion of the stock (Figure 10d).

3.4.2 Stochastic logistic growth

Consider now a renewable resource that grows logistically over time and is subject to stochastic perturbations, such as random events of individual mortality and reproduction in population dynamics (Lande et al., 2003). The evolution of the stock over time is now given by:

$$\mathrm{d}X_t = rX_t\left(1 - \frac{X_t}{K}\right)\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t \ , \tag{38}$$

where $\mathrm{d}W_t = W_{t+\mathrm{d}t} - W_t$ is the infinitesimal increment of a standard Wiener process W_t. That is, $\mathrm{d}W_t$ is a normally distributed random variable with mean zero and variance $\mathrm{d}t$. This random component is multiplied by the stock size X_t at time t, which means that the size of stochastic perturbations to the resource stock is proportional to the stock size. The system's susceptibility to stochastic influences is parametrized by σ.

Measuring causal responsibility in stochastic systems (Equation 34) requires specifying the interval $[\underline{x}, \bar{x}]$ centered around the realized system state x_t^a. Although the probabilities $p(X_t^a, \underline{x}, \bar{x})$ and $p(X_t, \underline{x}, \bar{x})$ change considerably for different interval widths, the relative difference between the probabilities and thus causal responsibility are not very sensitive to the interval width (Appendix Figure B.1). An agent's ex-post causal responsibility (Equation 34) for the actually realized system state x_t^a at time t depends on the specific, random realization of the stochastic process described by Equation (38).

Figure 11a shows one random realization (black line) for an action that modifies the initial value x_0. The expected values $\mathbb{E}[X_t^a]$ and $\mathbb{E}[X_t]$ of the corresponding actual and counterfactual process are slightly lower than in the deterministic case because stochastic fluctuations reduce the expected growth rate (Pindyck, 1984). The probability $p(X_t^a, 0.9x_t^a, 1.1x_t^a)$ of finding the process X_t^a within plus or minus ten percent of the realized system state x_t^a at time t is close to 1 initially because the variance of X_t^a is low initially. In the counterfactual case of not acting, the probability of finding the process X_t within the same interval is close to zero at first due to the low variance of X_t. As the variance of both processes increases over time, both probabilities tend to converge against each other and the agent's ex-post causal responsibility (Equation 34) decreases over time.

Figure 11b shows the agent's ex-ante causal responsibility (Equation 35) for the same action, which reveals the action's causal impact on the system in a representative manner. An agent's ex-ante causal responsibility for the resulting system state at time t is maximal at time 0 and subsequently decreases over time when modifying the initial value. Comparing with Figure 10a, it becomes apparent that the ex-ante responsibility for the stochastic system state is almost identical to causal responsibility for the deterministic system state. That is, the randomness inherent in a particular realization (Figure 11a) is smoothed over by averaging over a large number of realizations due to the law of large numbers.[24] Similar results are obtained for attractor and rate modifications (not shown here): an agent's ex-post causal responsibility depends on the particular realization of the system state, while the ex-ante causal responsibility resembles the corresponding deterministic case.

[24] This result is not a general property of the ex-ante causal responsibility and only holds for systems that can be described by a probability distribution of the exponential family, but not for e.g. heavy-tailed distributions.

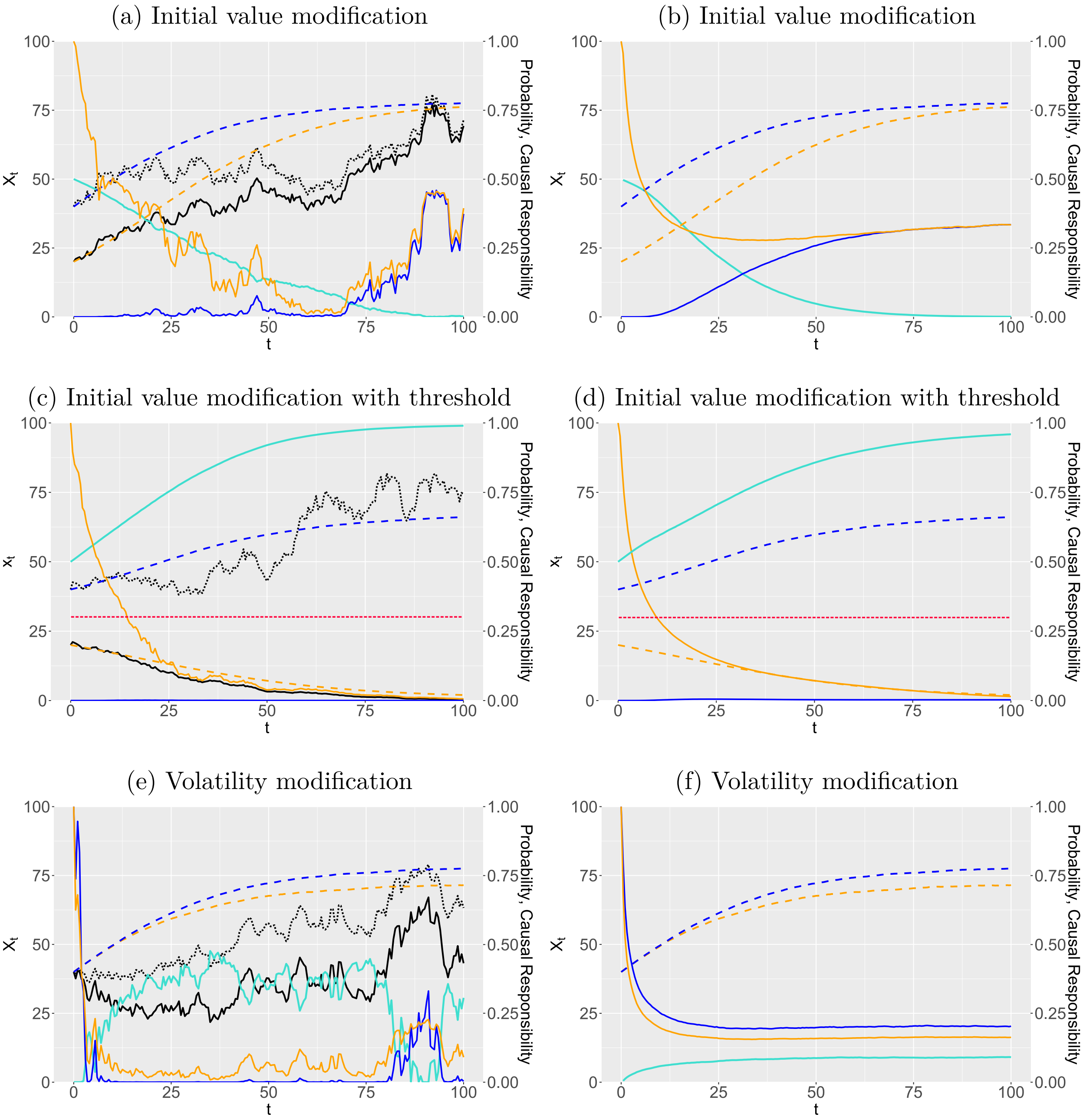


Figure 11: Ex-post and ex-ante causal responsibility for different actions under stochastic logistic stock dynamics with and without thresholds. Realized system state X_t^a (solid black), corresponding expected value $\mathbb{E}[X_t^a]$ (dashed orange) and probability $p(X_t^a, \underline{x}, \bar{x})$ (solid orange), counterfactual realization X_t (dotted black), corresponding expected value $\mathbb{E}[X_t]$ (dashed blue) and probability $p(X_t, \underline{x}, \bar{x})$ (solid blue), as well as ex-post causal responsibility $R(x_t^a, x_t)$ (Equation 34, panels a,c,e) and ex-ante causal responsibility $R^e(a)$ (Equation 35, panels b,d,f) (both solid turquoise) under stochastic logistic stock dynamics with and without thresholds (Equation 38 for a,b,e,f; Equation 40 for c and d). Parameter values: $r = 0.05, K = 80, x_0 = 40, \sigma = 0.05, \underline{x} = 0.9x_t^a, \bar{x} = 1.1x_t^a$ in a-f, $x_0^a = 20$ in a-d, $V = 30$ in c and d (dashed red), $\sigma^a = 0.1$ in e and f.

Further, the agent can modify the system's susceptibility to stochastic shocks by changing the value of σ (Figure 11e). Increasing σ leads to a higher variance of the process X_t^a (Equation 38) and thus a lower probability of finding it relatively close to its expected value. Hence, for realized system states close to $\mathbb{E}[X_t^a]$ the probability due to action is lower than the probability in the counterfactual case of not acting. Conversely, larger deviations of X_t from the expected system state become more likely by increasing σ. Hence, an agent's ex-post causal responsibility for the realized system state is larger the farther X_t deviates from $\mathbb{E}[X_t^a]$ when increasing σ.

An agent's ex-ante causal responsibility for the resulting system state is zero initially and subsequently increases over time for actions that increase σ (Figure 11f). As the variance approaches its stationary level, causal responsibility converges against its maximum level over time. Although on average the probability due to action is lower than the counterfactual probability, the ex-ante causal responsibility is positive due to possible realizations far from the expected value.

Figure 12 depicts the case of a natural resource with economically optimal harvesting that maximizes discounted net surplus for isoelastic demand and marginal cost functions.[25] The agent's ex-post causal responsibility increases over time as the difference between the exploited stock and the counterfactual stock without extraction increases and converges against its maximum level (Figure 12a). In general, causal responsibility is relatively high when choosing an optimal control strategy, because the system is exploited to a strong degree (Figure 12a and 12b). Under certain economic or biological conditions, such as a high discount rate, it may be economically optimal to drive the stock to extinction, for which the agent is then fully causally responsible (Figure 12c and 12d).

Finally, consider a logistically growing renewable resource that is subject to stochastic perturbations and exhibits critical depensation. The stock dynamics are given by:

$$\mathrm{d}X_t = rX_t\left(1-\frac{X_t}{K}\right)\left(\frac{X_t}{V}-1\right)\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t\ . \tag{40}$$

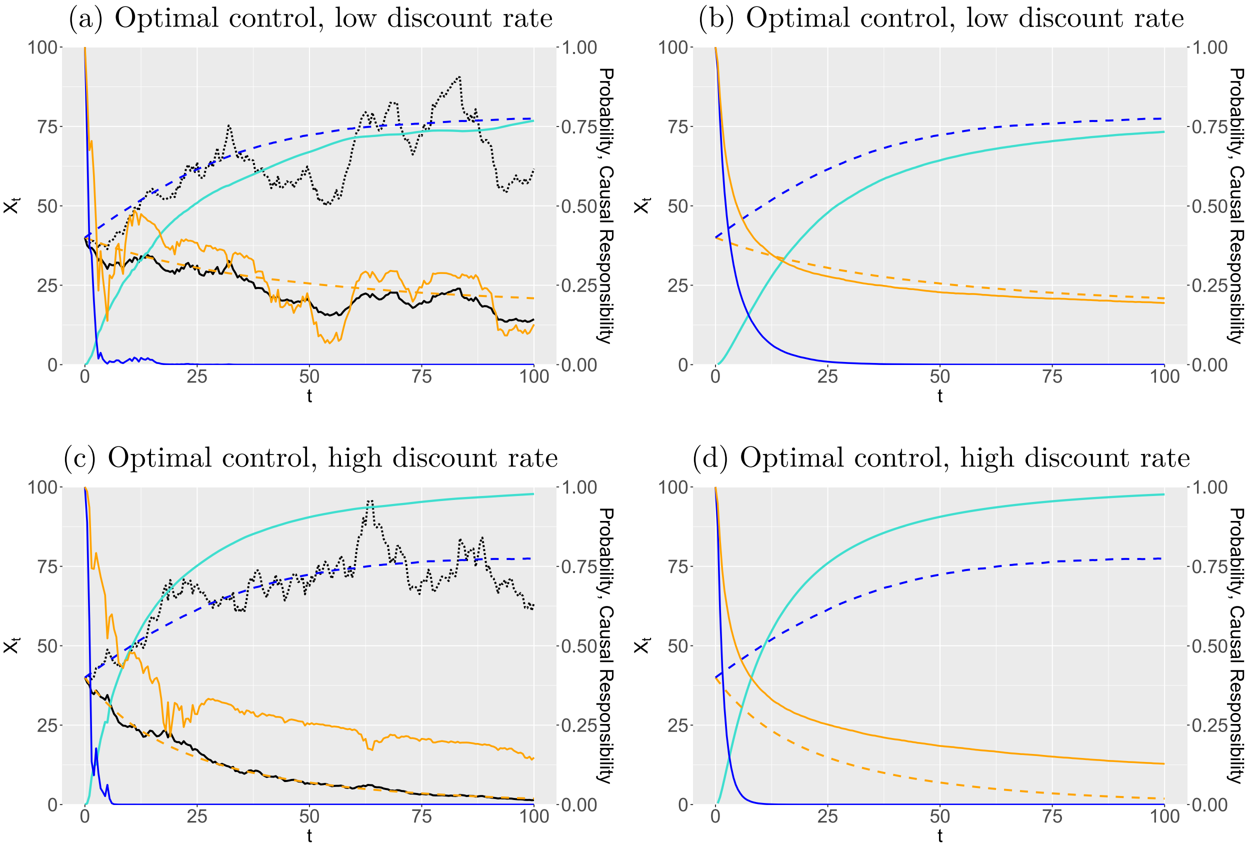
An agent's ex-post causal responsibility for an action that decreases the initial value below the threshold value V depends less on the particular realization than when not crossing

[25] The optimal extraction rule $h^*(X_t)$ in this case is given by (Pindyck, 1984):

$$h^*(X_t) = bX_t\left\langle c + \left\{2b^2 + 2b\left[b^2 + c\left(r+\delta-\sigma^2\right)^2\right]^{\frac{1}{2}}\right\}\Big/\left(r+\delta-\sigma^2\right)^2\right\rangle^{-\frac{1}{2}}, \tag{39}$$

with isoleastic demand $q(p) = bp^{-\eta}$ with $\eta = 1/2$ and isoelastic marginal cost $c(X_t) = cX_t^{-\gamma}$ with $\gamma = 2$ and discount rate δ.

the threshold (Figure 11c).[26] In particular, the action's degree of sufficiency approaches 1 over time, whereas its degree of necessity is close to 1 throughout, although the probability $p(X_t^a, \underline{x}, \bar{x})$ decreases sharply over time as the variance decreases.



(a) Optimal control, low discount rate

(b) Optimal control, low discount rate

(c) Optimal control, high discount rate

(d) Optimal control, high discount rate

Figure 12: Causal responsibility in stochastic systems for economically optimal control. Realized system state X_t^a (solid black), corresponding expected value $\mathbb{E}[X_t^a]$ (dashed orange) and probability $p(X_t^a, \underline{x}, \bar{x})$ (solid orange), counterfactual realization X_t (dotted black), corresponding expected value $\mathbb{E}[X_t]$ (dashed blue) and probability $p(X_t, \underline{x}, \bar{x})$ (solid blue), as well as ex-post causal responsibility $R(X_t^a, X_t)$ (Equation 34, panel a) and ex-ante causal responsibility $R^e(a)$ (Equation 35, panel b) (both solid turquoise) under stochastic logistic stock dynamics (Equation 38) and economically optimal harvesting $h^*(X_t)$ (Equation 39). Parameter values: $r = 0.05, K = 80, x_0 = 40, \sigma = 0.05, b = 1, c = 1, \underline{x} = 0.9x_t^a, \bar{x} = 1.1x_t^a$ in a-d, $\delta = 0.03$ in a and b, $\delta = 0.1$ in c and d.

Similar to the deterministic case (Figure 10d), an agent's ex-ante causal responsibility for the resulting system state when taking an action that decreases the initial value below the critical threshold is increasing and converges to its maximum value over time (Figure 11d).

[26] The magnitude of this effect depends on the parameter values. Here, proportional stochastic perturbations to the system state are very small because the system state itself is very small.

It is slightly lower than causal responsibility in the deterministic case due to the (unlikely) possibility that the counterfactual system state decreases below the threshold value, or that X_t^a increases above the threshold, due to stochastic perturbations.

3.4.3 General results and conjectures

Beyond specific example systems and actions, we now formulate general results for causal responsibility in dynamical systems. These are deduced from the insights gained from an encompassing set of examples, rather than being derived analytically from Section 3.2 in an elementary manner. In that sense, they are conjectures, yet well-founded and reasoned.

While these results are fairly general, they only apply to systems that have at least one locally stable non-trivial equilibrium and do not exhibit cyclical or chaotic behavior. This may exclude certain parameter values and actions even in the examples presented (such as large values of r in the logistic growth model, which give rise to chaotic behavior). As throughout the entire analysis, we remain in the setting described in Section 3.2: a single action's consequences unfold under (probabilistically) known circumstances.

We focus on how an agent's ex-ante causal responsibility develops over the long run. One essential result is that causal responsibility may increase or decrease over time, depending on the system and action type. More specifically, causal responsibility may either vanish asymptotically over time, or it may converge to a finite, constant level.

Table 2: Long-run development of ex-ante causal responsibility, depending on the type of system and on the action type

Action type	**D**	**DT**	**S**	**ST**
Initial value (x_0)	vanishing	vanishing or lasting	vanishing	vanishing or lasting
Attractor (K)	lasting	vanishing or lasting	lasting	vanishing or lasting
Rate (r)	vanishing	vanishing	vanishing	vanishing
Volatility (σ)	–	–	lasting	lasting
Control strategy (h)	lasting	lasting	lasting	lasting

D=deterministic systems without thresholds, **DT**=deterministic systems with thresholds, **S**=stochastic systems without thresholds, **ST**=stochastic systems with thresholds

For systems without thresholds, the long-run development of causal responsibility is determined by the action type: initial value and rate modifications entail vanishing causal responsibility, whereas attractor and volatility modifications as well as the choice of any control strategy entail lasting causal responsibility. For systems with thresholds, the long-run development of causal responsibility for some action types also depends on other factors.

For initial value modifications, causal responsibility is vanishing if the action does not cause the system to cross the threshold, whereas it is lasting if it does. For attractor modifications, causal responsibility is vanishing if the system is initially below its threshold, and lasting if it is above. Table 2 summarizes these results.

To quantitatively describe the temporal extent of causal responsibility, we introduce a significance threshold $\overline{R}$, which represents the minimum level of causal responsibility below which an action's causal impact is deemed negligible. The actual value of $\overline{R}$ is not an inherent property of the system, but reflects the (risk) preferences of society. It follows that causal responsibility may be limited in time by falling below this threshold. Formally, the time period $\mathcal{T}^{\text{sig}}(a)$ during which an action a's causal impact on the system is significant is defined by:

$$\mathcal{T}^{\text{sig}}(a) := \left\{t \,|\, R^e\big(a]\big) \geq \overline{R}\right\} \ . \tag{41}$$

For cases of vanishing causal responsibility there exists some $T^{\max}(a) := \sup \mathcal{T}^{\text{sig}}(a)$. After this point in time action a no longer exerts a significant causal influence on the system (Figure 13).[27] It describes the maximum temporal extent of an action's causal efficacy on the system. Likewise, there exists a minimum time $T^{\min}(a) := \inf \mathcal{T}^{\text{sig}}(a)$ before which the action a has no significant causal efficacy on the system. This time lag, which may be zero, between the time of action and when the action's consequences begin to take a significant effect is well-known in the context of monetary policy (e.g. Friedman, 1961) but is relevant for policy-making more generally.

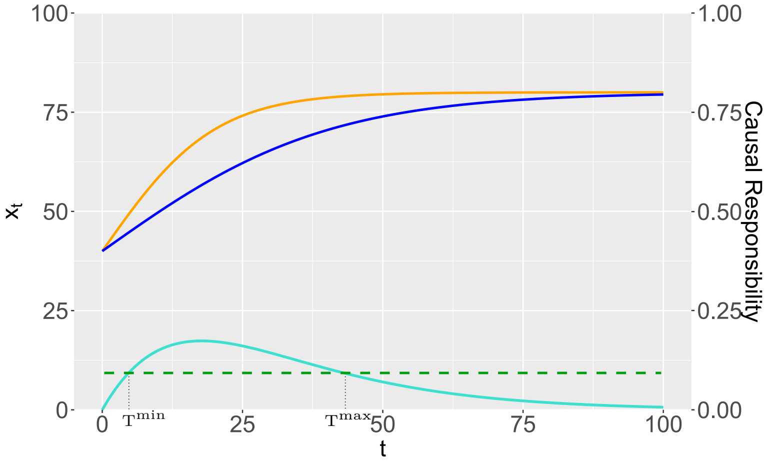


Figure 13: Time period during which causal responsibility is significant. Actual system state X_t^a (solid orange), counterfactual system state X_t (solid blue) and causal responsibility $R(X_t^a, X_t)$ (Equation 33) (solid turquoise) under deterministic logistic stock dynamics (Equation 36) with significance threshold $\overline{R}$ (dashed green). Parameter values: $r = 0.05, r^a = 0.1, K = 80, x_0 = 40, \overline{R} = 0.1$.

[27] Formally, if causal responsibility is lasting, $T^{\max}(a)$ is not defined, but infinite.

3.5 Relevance

Our results show that the time of occurrence of a system state is crucial for the extent of causal responsibility. The underlying fundamental reason is that the relationship between cause and effect may change over time. This aspect is neglected when one performs a (quasi-) static assessment of causality in a dynamical system. Our concept is relevant whenever the action's consequences dynamically unfold in a non-trivial way because it explicitly captures this aspect. In particular, this may be relevant in the following instances.

3.5.1 Attribution and impact assessment

Obviously, our concept of causal responsibility can be used to attribute an observed system state to its causes (ex-post) and to assess the expected causal efficacy of different actions (ex-ante). This is relevant for formulating feasible management goals, assessing the effectiveness of management actions for given goals, appropriately setting economic incentives, and judging the quality of management actions as a basis for reward or punishment.

If one thinks of actions as policy measures, our concept allows an – ex-ante or ex-post – assessment of their effectiveness to reach a given target system state. The assessment is in terms of a single number, which means it could be used as an indicator of effectiveness. Examples include policies which aim at reaching a predefined system state, such as an inflation target, full employment, a public health target (e.g., vaccination rates), or "good status" of freshwater bodies (defined through threshold values).

If one asks whether a given agent is to blame or praise for the state of a dynamical system, our concept allows an ex-post attribution of the system state to the agent's action and natural dynamics. For example, our concept quantitatively measures to what extent a mining company's discharge of pollutants into a river has caused the subsequent collapse of a fish stock.

3.5.2 Liability

Our concept is relevant for the design of strict[28] liability regulations when an agent's action subsequently entails a damage to another person. In particular, suppose the damage is determined by the actually realized system state. If the agent has (partially) caused this system state she is liable, in principle, for compensation. In the law-and-economics literature on liability, different institutional designs have been discussed in terms of whether they can

[28] Strict liability follows the logic of consequentialism. Hence, causation is at its core, in contrast to negligence liability (Epstein, 1973).

establish appropriate incentives for an efficient allocation (Shavell, 1987; Pitchford, 1995; Alberini & Austin, 2002; Boomhower, 2019). In contrast to designing liability regulations solely on grounds of efficiency, one may also design liability in proportion to the agent's causal responsibility for the damaging system state, which is both efficient and in line with generally accepted principles of causation (Baumgärtner & Quaas, 2021). More precisely, liability in proportion to causal responsibility means that the agent owes compensation of that fraction of the damage for which she is causally responsible.

Our concept captures how the causal relationship between the damage and the agent's action changes over time, which is relevant when designing liability regulations in dynamical systems in proportion to causal responsibility. First, if a damage occurs at a point in time subsequent to the agent's action, the agent's degree of causation of the damage depends not on the actual and counterfactual system state at the time of action, but on the actual and counterfactual system state at the time of damage. Accordingly, the extent of the agent's liability crucially depends on the time at which the damage occurs.

Second, if a damage occurs over an extended period of time, the agent's degree of causation of the damage may be different at each point in time. Hence, at each point in time during the damage period the agent is liable for compensation of that fraction of the damage for which she is causally responsible at that time. As this fraction is not necessarily constant over time, it needs to be factored in at each point in time when assessing the agent's liability for the total damage over the entire time period.

3.5.3 Normative responsibility

The concept of responsibility, in general, has different layers of meaning (Baumgärtner et al., 2018, Sec. 3.1). We have so far focused on the elementary layer of causal responsibility, which is purely descriptive. We now turn to normative responsibility, which is about how one *should* act. In particular, we discuss the implications of causal responsibility for normative responsibility in dynamical systems.

Our understanding of normative responsibility is founded on consequentialist ethics, according to which actions are judged based on their consequences.[29] In a dynamical system, an agent's normative responsibility is to effectuate a future desired system state, or to avoid an undesired one, by choosing at time 0 a suitable action from the actions at her disposal. For example, the agent's normative responsibility may be to see to it that a natural resource is not exhausted.

Generally, the extent of normative responsibility may be limited due to several reasons

[29] This is opposed to deontological ethics, according to which actions are considered morally right or wrong irrespective of their consequences (Alexander & Moore, 2021).

(Baumgärtner et al., 2018, Sec. 4.4). One important reason is the agent's limited causal responsibility, that is, the agent's limited ability to effectuate or avoid a normatively specified future system state. This fundamental limit has been introduced by Immanuel Kant (cf. Stern, 2004) and is known in modern ethics as the Ought-Implies-Can-Principle (Van Inwagen, 1978; Griffin, 1992): one can only be obliged to do what one is able to do. In other words, being able to effectuate a particular system state is a necessary condition for bearing normative responsibility for it.

As discussed in Section 3.4.3, an agent's causal efficacy when taking a particular action a may be limited in time. Accordingly, the agent's normative responsibility may also be limited in time. In particular, if an agent's causal responsibility for the system state at time t is below the significance threshold $\overline{R}$ for *any* action at her disposal, the agent cannot be normatively responsible for the system state at that time. That is, the temporal extent of an agent's normative responsibility cannot extend beyond the time period during which the agent's causal impact on the system is significant when considering all possible actions. For actions that entail vanishing causal responsibility (see Table 2), the agent's normative responsibility for future system states is therefore limited by the largest $T^{\max}(a)$ of all actions at the agent's disposal. If there exists a time lag between the time of action and when the action's consequences begin to take a significant effect, the agent cannot be normatively responsible for system states before the smallest $T^{\min}(a)$ when considering all actions at her disposal.

The Ought-Implies-Can-Principle thus limits the temporal extent of an agent's normative responsibility. These limits need to be respected when specifying an agent's normative responsibility.

3.6 Discussion and conclusion

We have developed a novel measure of an agent's degree of causal responsibility for the state of dynamical systems founded on the agent's action's degree of necessity and sufficiency for the system state. Going beyond existing quantitative measures of the degree of causation of a given outcome, our concept captures the varying strength of causal relationships over time and can be applied in deterministic and stochastic systems for both discrete and continuous conceptions of the system state. We have shown that the extent and trajectory of causal responsibility over time vary substantially both across different types of systems for identical actions and across different types of actions within the same system. For given type of system and action, the extent of causal responsibility is determined – by definition – by the time at which a particular system state occurs.

We have applied this general measure of causal responsibility to different stylized actions in a number of simple example systems. Applying our concept to more complex actions in real-world systems requires good system knowledge formalized in a dynamic model. For many systems, such detailed knowledge in the form of a model might not yet be available, for instance due to limited data. Still, the practice of attributing extreme weather events to climate change (Allen, 2003; Stott et al., 2004; Otto, 2017) exemplifies that it is possible to make robust counterfactual predictions despite highly complex system dynamics.

Our measure of causal responsibility is independent of any norm about how the system state ought to be or what action ought to be taken. While causation itself is purely descriptive, ascribing causality to an agent does carry some normative content about how the attribution should be done. For instance, it needs to be specified what knowledge about the action's consequences can reasonably be expected of the agent. Here, we assumed that the agent is fully aware of the state of probabilistic knowledge available at the time of action. Furthermore, when using a counterfactual conception of causation, it needs to be specified against which reference action the action is compared. Here, we took not acting as the reference action. This reflects the conventional view that acting in a dynamical system means interfering with the natural dynamics and not acting being the default.

We deliberately restricted our analysis to systems with a single state variable, since a single measure of causal responsibility for a multi-dimensional system state would require some form of aggregation. It is well-known that such aggregation cannot be done in a descriptive and value-free manner.

To focus on the dynamic aspect of causation in stochastic dynamical systems, we analyzed an agent's degree of causal responsibility for the realized system state at a particular point in time. An obvious extension would be to assess an agent's degree of causal responsibility for the trajectory of the system state over some time interval. For instance, one building block of such a measure could be the L^1-norm of the realized and the counterfactual process, indicating how much the action changes the continuous trajectory of the system state over this time interval (Krysiak, 2011). By such an aggregation, one would gain insight into the action's overall impact over an extended time interval, indicated by a single number. Yet, one would lose more detailed information about the degree of causation at each point in time.

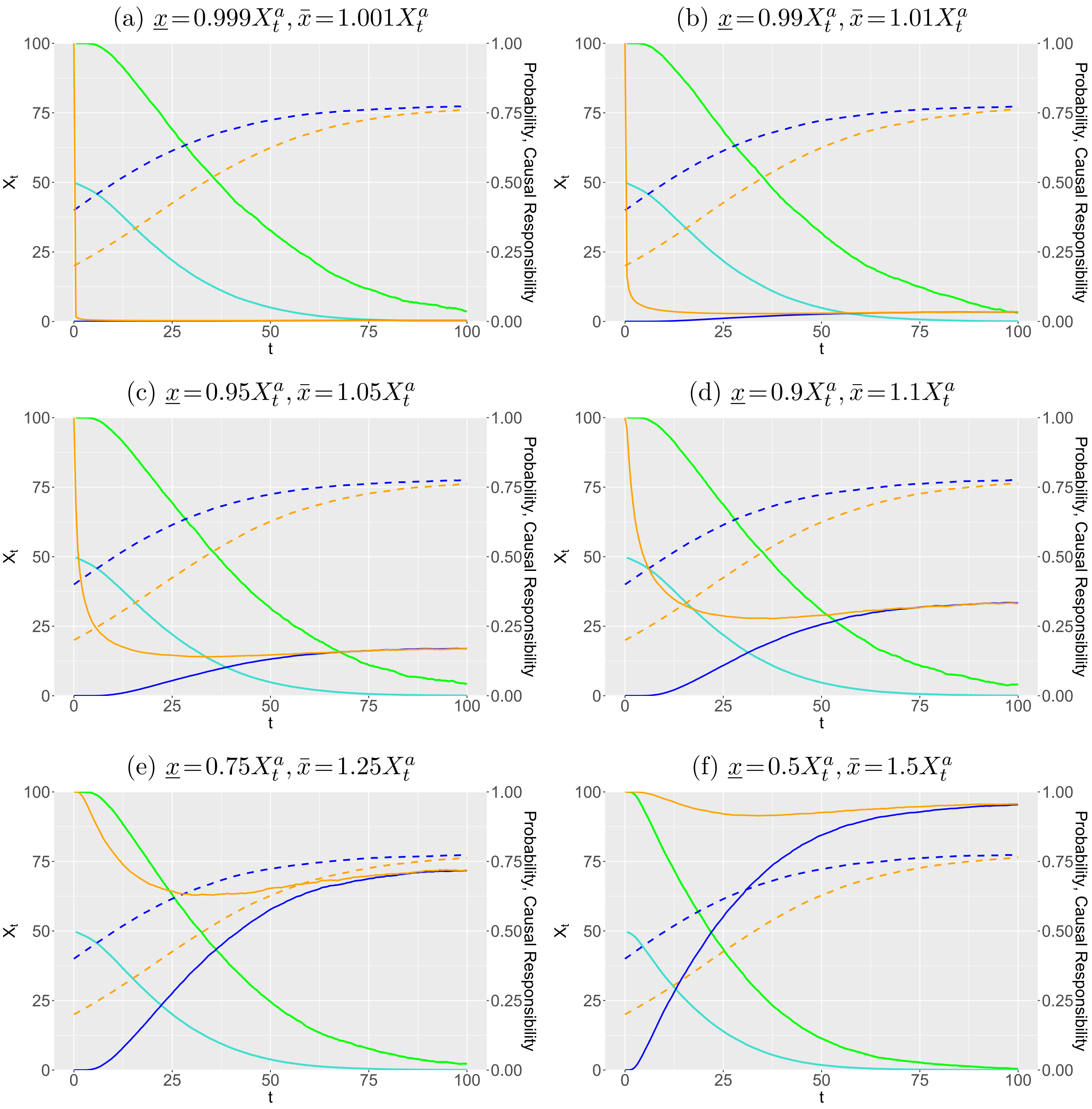
Another restriction of our analysis is the single-agent setup, which allows a clear focus on the properties of causal responsibility in stochastic dynamical systems. Of course, in most relevant problems, many agents are involved. For the case of multiple agents acting sequentially with complete knowledge, each agent's causal responsibility can be assessed by applying our concept, with the system dynamics as determined by previous actions forming the counterfactual reference. When multiple agents act simultaneously or with incomplete knowledge, one needs a more complicated scheme to attribute the jointly caused outcome to each agent individually. Concepts for measuring causation in such a multi-agent setting exist (e.g. Chockler & Halpern, 2004; Braham & van Hees, 2009; Mittelstaedt & Baumgärtner, 2023), but involve other strong simplifications, such as omission of dynamics, stochasticity, management, or continuity of the system state. Generalizing our concept to a multi-setting is a considerable challenge for future research.

In conclusion, our measure of causal responsibility is relevant whenever an action's consequences dynamically unfold in a non-trivial way. It can be used to attribute a realized system state to its causes, to quantitatively assess the effectiveness of management actions and policies over time, to design liability regulations that are both in line with causality and economically efficient, and to delineate the temporal scope of an agent's normative responsibility.

Acknowledgements

We are grateful to Benjamin Blanz, Anne-Sophie Crépin, Carsten Dormann, Nora Felber, Hermann Held, Daniel Heyen, Frank Krysiak, Christian Mittelstaedt, Frikk Nesje, Daniele Rinaldo and Theresa Voss for discussion and comments, and to the German Federal Ministry of Education and Research (BMBF) for financial support under grant no. 01LC1826C.

Appendix B Additional Simulations



(a) $\underline{x}=0.999X_t^a, \bar{x}=1.001X_t^a$

(b) $\underline{x}=0.99X_t^a, \bar{x}=1.01X_t^a$

(c) $\underline{x}=0.95X_t^a, \bar{x}=1.05X_t^a$

(d) $\underline{x}=0.9X_t^a, \bar{x}=1.1X_t^a$

(e) $\underline{x}=0.75X_t^a, \bar{x}=1.25X_t^a$

(f) $\underline{x}=0.5X_t^a, \bar{x}=1.5X_t^a$

Figure B.1: Ex-ante causal responsibility and degree of necessity for different interval widths. Actual expected value $\mathbb{E}[X_t^a]$ and probability $p(X_t^a, \underline{x}, \bar{x})$, counterfactual expected value $\mathbb{E}[X_t]$ and probability $p(X_t, \underline{x}, \bar{x})$, as well as ex-ante causal responsibility $R^e(a)$ (Equation 35) and degree of necessity DN (first factor in Equation 34) under stochastic logistic stock dynamics (Equation 38). Parameter values: $r=0.05, K=80, x_0=40, \sigma=0.05, x_0^a=20, \underline{x}$ and $\bar{x}$ given in panel captions.

Chapter 4

Attribution of fish stock collapse to overfishing and climate change

This chapter was written with Christian Möllmann, Martin Quaas, and Stefan Baumgärtner.*

This chapter has been submitted for publication to *Nature Communications Earth & Environment* and is currently under review.

Abstract: The long-standing debate about whether fishing pressure or environmental factors are to blame for fish stock collapses and their severe ecological, social, and economic consequences has largely been led with qualitative arguments based on anecdotal evidence. Here, we propose a new method to give a nuanced quantitative answer to this question through a model-based causal attribution procedure. We apply this to the case of the recently collapsed Western Baltic cod stock using a stochastic cusp model, ICES stock assessment data, and an attribution scheme based on the Shapley value. We quantify the respective contributions of overfishing and climate change to causing the collapse by assessing the extent to which they have increased its likelihood relative to counterfactual scenarios in which either one or both factors are absent. We find that the extent to which overfishing has caused the collapse was 75%, climate change 18%, and other factors 7% – with considerable uncertainty due to limited data quality. Our generic model-based causal attribution procedure is very general and can be used to quantify human impacts in (mis-)managed ecosystems for which a stochastic model and sufficient data exist.

*Author contributions:
Michael Stecher: conceptualization (equal), formal analysis (lead), methodology (lead), software (lead), visualization (lead), writing – original draft (lead), writing – review & editing (lead), **Christian Möllmann:** conceptualization (equal), funding acquisition (equal), methodology (supporting), writing – review & editing (supporting), **Martin Quaas:** conceptualization (equal), funding acquisition (equal), methodology (supporting), writing – review & editing (supporting), **Stefan Baumgärtner:** conceptualization (equal), funding acquisition (equal), methodology (supporting), supervision (lead), writing – review & editing (supporting)

4.1 Introduction

Many fish stocks around the world have experienced dramatic declines of their biomass and productivity, with often disastrous ecological, economic and social consequences (Cook et al., 1997; Myers et al., 1997; Casey & Myers, 1998; Hutchings, 2000; Jiao, 2009). The abrupt nature of these stock collapses and the common failure to recover from them are thought to originate from regime shifts in the corresponding ecosystems (deYoung et al., 2008; Vert-Pre et al., 2013; Möllmann et al., 2015). What exactly has caused these stock collapses or the underlying regime shifts has been the subject of a long-standing and highly contentious debate among scientists (Pershing et al., 2015; Palmer et al., 2016; Swain et al., 2016; Pershing et al., 2016; Brander, 2018; Froese et al., 2022) and stakeholders. While unsustainably high fishing pressure is the obvious culprit for some, others blame environmental factors, most notably climate change, for these events. These conclusions are typically based on visual comparisons, correlations, or ad-hoc measures (Beaugrand et al., 2022) of observed deterministic trends in fishing mortality, ocean temperatures, and stock size. What has been missing from this debate is a nuanced assessment of the degree to which overfishing and climate change have actually caused a particular stock collapse, taking into account the stochastic nature of stock dynamics.

Atlantic cod (*Gadus morhua*) has been one of the most important and most heavily exploited fish species and many cod stocks across the North Atlantic have experienced dramatic collapses (Sguotti et al., 2019). The Western Baltic cod (WBC) stock is a relatively minor, yet socio-culturally important (Döring et al., 2020) cod population resident in the Western Baltic Sea, a small part of the Baltic Sea characterized by above-average warming (Dutheil et al., 2021). Its biomass has declined by roughly 90% and catches have fallen by more than 95% since 1997 (ICES, 2021). The WBC stock was able to recover from an earlier collapse in the mid-1980s, but has not recovered from its recent collapse in the mid-2000s (Figure 14a). This can be explained by the stock having crossed a tipping point into a regime characterized by low stock size and low productivity, in which it has subsequently stabilized (Möllmann et al., 2021). The combined effect of overfishing and climate change has been identified as the cause of this shift (Möllmann et al., 2021), but their respective contributions have not been assessed quantitatively so far.

Here, we quantitatively measure the extent to which fishing pressure (Figure 14d), ocean warming (Figure 14b), and other factors have caused the recent collapse of the WBC stock. To this end, we develop a systematic attribution procedure that combines three elements. We calibrate (i) a stochastic cusp model (Cobb & Watson, 1980; Cobb et al., 1983; Grasman et al., 2010) with (ii) stock assessment data for WBC (ICES, 2021).

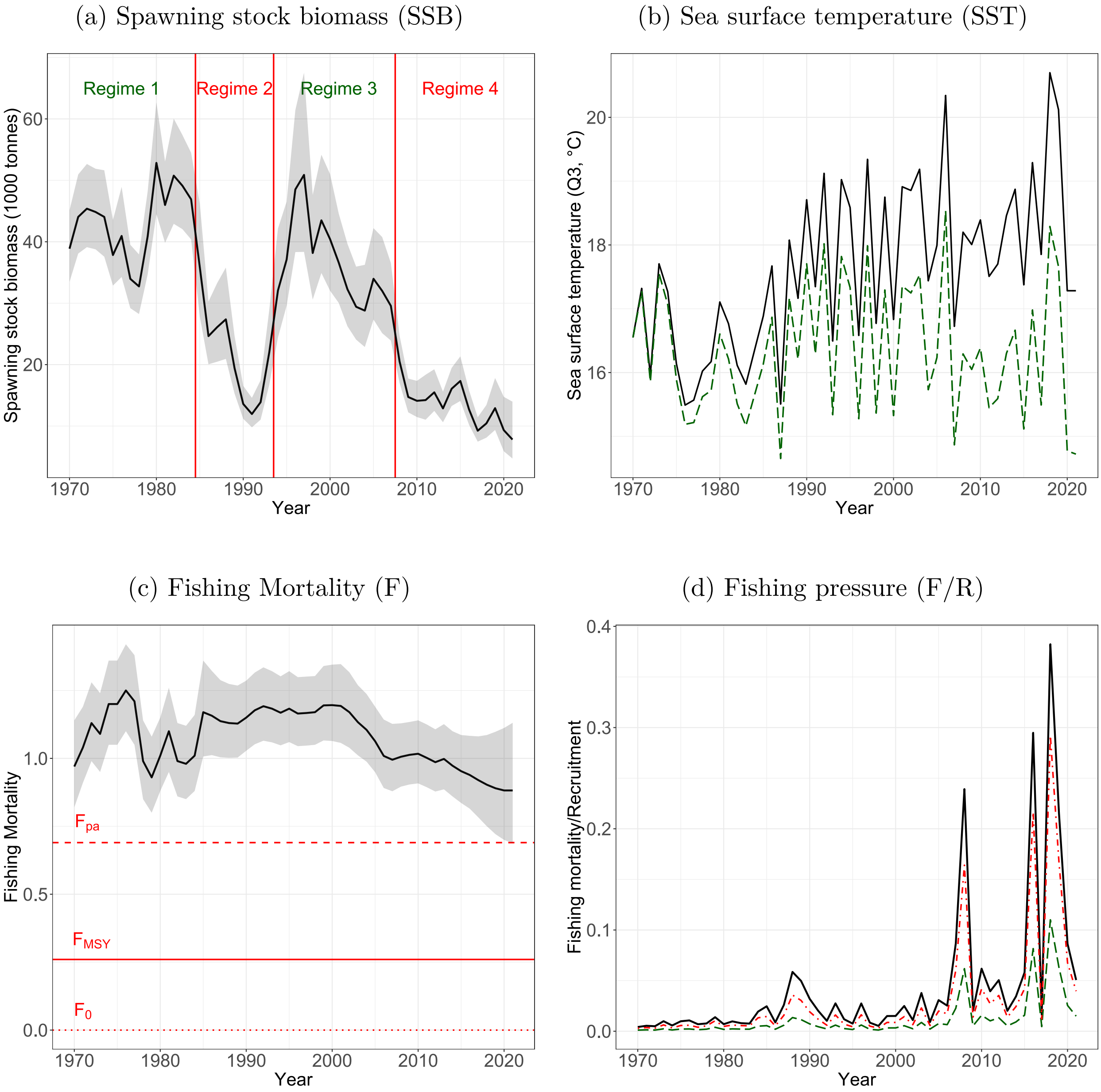


Figure 14: Input data used for model calibration and counterfactual simulations. Black solid lines denote factual data; grey shaded areas in (a) and (c) denote 95% confidence intervals; red vertical lines in (a) separate regimes identified by Möllmann et al. (2021); red horizontal lines in (c) correspond to zero (dotted), maximum sustainable yield (solid) and precautionary (dashed) reference points for fishing mortality; green dashed line in (b) denotes detrended SST; red dot-dashed line in (d) denotes F_{pa}/R, green dashed line in (d) denotes F_{MSY}/R. Data: see Methods.

To divide the combined effect of overfishing and climate change on the likelihood of collapse into the respective contributions of these factors, we employ (iii) the attribution scheme of

Mittelstaedt & Baumgärtner (2023) for attributing regime shifts in stochastic systems to individual actors. This scheme, which is based on the Shapley value (Shapley, 1953), considers the difference in likelihood of collapse that overfishing and climate change, respectively, make relative to counterfactual scenarios in which either fishing pressure, or ocean warming, or both are absent. Overall, our attribution procedure is founded on a counterfactual, probabilistic, ex-post conception of causation in stochastic systems (Stecher & Baumgärtner, 2022a). We find that the extent to which overfishing has caused the collapse was 75%, climate change 18%, and other factors 7% – with considerable uncertainty due to limited data quality (see Discussion).

Beyond WBC and other collapsed fish stocks, our systematic attribution procedure can be used to quantify the respective contributions of multiple factors to a given ecosystem state. In particular, it can be used to quantify human impacts in (mis-)managed ecosystems. The procedure is generally based on three elements: (i) model, (ii) data and (iii) attribution scheme. According to the system under study and research objective, one may choose a different model, data set, or attribution scheme.

Quantitative causal attribution has been established in climate science where specific climatic events (e.g., heatwaves or floods) are attributed to anthropogenic climate change (Sippel et al., 2020). Specifically, climate model simulations with and without anthropogenic greenhouse gas emissions make it possible to compute the fraction of the risk of a particular event that is attributable to greenhouse gas emissions (Allen, 2003; Stott et al., 2004; Otto, 2017). Our procedure allows, for the first time, a similar attribution of biodiversity loss to several potential causes. While the procedure is similar, the main difference is the number of causes to which an event is attributed: climate attribution considers a single factor (anthropogenic climate change) other than natural variability, whereas we consider more than one (fishing pressure and ocean warming). Our main methodological innovation for this generalization is the use of an attribution scheme for multiple factors (element iii).

4.2 Results

Due to the inherent stochasticity of many ecological processes, such as stock recruitment and mortality, regime shifts in marine ecosystems can be regarded as stochastic events that may occur with a certain probability determined by anthropogenic and environmental factors (Hsieh et al., 2005). Attributing a regime shift to its causes thus needs to be based on probabilistic information. The basic idea of probabilistic attribution is that a factor's contribution to causing a specific event is given by the relative increase in the likelihood of the event in the presence of this factor (Hannart et al., 2016). To determine how the likelihood of a shift to a low-productivity regime and the ensuing collapse of WBC has changed in the presence of overfishing and climate change, we simulate different counterfactual reference scenarios in which either fishing pressure, or ocean warming, or both are absent. To this end, we first calibrate a stochastic cusp model (SCM) to WBC using data on spawning stock biomass (SSB), recruitment (R), fishing mortality (F), and sea surface temperature (SST) (see Methods).

SCM is an approach to model how multiple factors interact in facilitating regime shifts in a stochastic system (Cobb & Watson, 1980; Cobb et al., 1983; Grasman et al., 2010). In particular, SCM describes how the response of a state variable to a so-called asymmetry factor changes from continuous to discontinuous depending on a second, so-called bifurcation factor (see Methods). We follow existing applications of SCM to cod stocks (Sguotti et al., 2019; Möllmann et al., 2021) by modeling the dynamics of the state variable as a function of SSB and fitting the asymmetry and bifurcation factors as functions of fishing pressure and ocean warming, respectively. We use fishing mortality scaled to recruitment (F/R) as a modified measure of fishing pressure that better explains SSB dynamics (Möllmann et al., 2021).

Stability properties of WBC

SCM provides insight into the stability properties of the stock. Depending on the level of fishing pressure and ocean warming, SSB may exhibit either a single stable equilibrium or two locally stable equilibria. This gives rise to two different possibilities of how a regime shift may occur. The first possibility is that stock dynamics are bistable and the stock crosses from the domain of attraction of one locally stable equilibrium into that of the other one. The boundary between the two domains of attraction represents a threshold value of SSB – when crossed, a regime shift occurs. The second possibility is that one of the locally stable equilibria does no longer exist for a certain level of fishing pressure and ocean warming. A regime shift occurs when the stock necessarily converges to the remaining, single stable equilibrium, which is known as a “critical transition” (Scheffer, 2009).

The calibrated SCM suggests that the dynamics of WBC were mostly bistable during the time period analyzed (Figure 15). The threshold value separating the alternative locally stable equilibria tends to increase with higher fishing pressure and SST. Since 2007, the stock dynamics have increasingly stabilized in a low-productivity regime – indicated by the absence of a second, high-productivity equilibrium – due to the combined effect of progressing climate change and a level of fishing pressure that is excessive for the diminished productivity of the stock. This makes a recovery of the stock to previous biomass levels increasingly unlikely, indicating hysteresis, i.e., a low degree of reversibility.

The instantaneous probability, in year t, of a regime shift to occur in year $t+1$ is given by the probability of SSB crossing its threshold value in year $t+1$ and denoted by p_{rs}. It is zero when, in two consecutive years, there is only one and the same equilibrium, as then there is no alternative regime to which SSB could shift. A regime shift is certain to occur if the equilibrium in year t does no longer exist in year $t+1$ (see Methods). For factual levels of fishing pressure and SST (Figures 14b and 14d, black solid lines) in the presence of fishing (F) and climate change (C), the model predicts the recent shift of WBC to its current, low-productivity regime to occur with a probability of $p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}) = 100\%$ in 2007 (Figure 15, dashed blue line). That is, the model suggests that a critical transition to a low-productivity regime and the ensuing collapse of the stock was inevitable given the observed levels of fishing pressure and ocean warming because the prevailing high-productivity equilibrium did no longer exist.

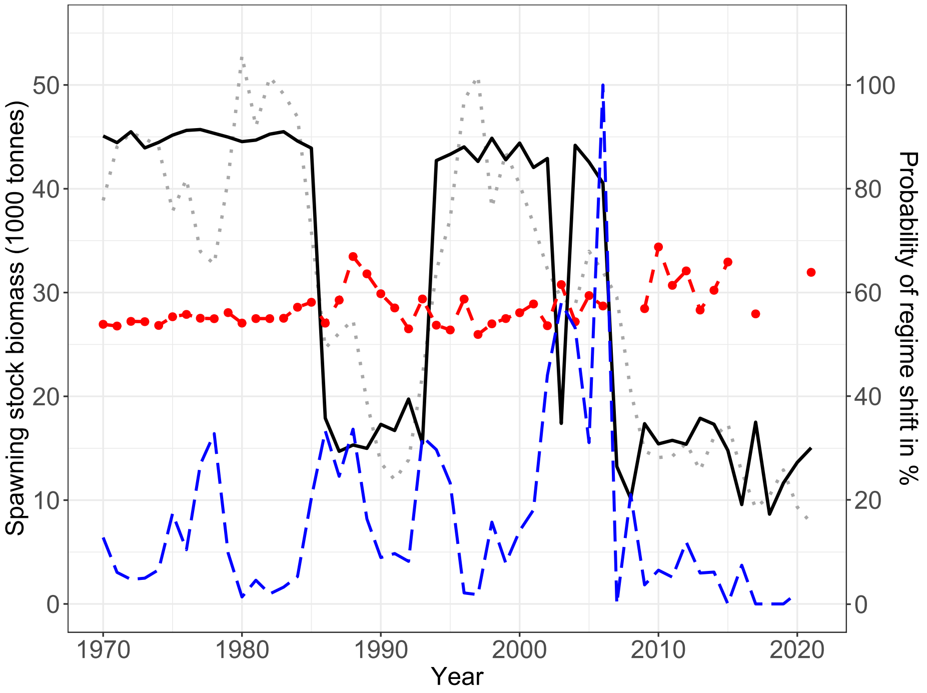


Figure 15: Model predictions for spawning stock biomass, threshold value and probability of regime shift. Predicted SSB (solid black line), factual SSB (dotted gray line), threshold value (dot-dashed red line) and instantaneous probability of regime shift $p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C})$ (dashed blue line) over time.

Simulating counterfactual scenarios

In the next step of our causal attribution procedure, we use the calibrated SCM to simulate three different counterfactual scenarios in which, throughout the entire time period, either fishing pressure is zero, or there is no ocean warming, or both factors are absent. This allows us to assess the difference that these factors made to the likelihood of a regime shift. We denote the probability of regime shift in the presence of both factors as $p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C})$ and as $p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0)$ in their absence. The probabilities when only one of the factors is present and the other absent are denoted by $p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}_0)$ and $p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C})$.

In the baseline counterfactual scenario both fishing and climate change are absent. As a consequence, the stock remains in the high-productivity regime. Although a low-productivity equilibrium exists under the given conditions, the probability of a shift to the low-productivity regime remains low throughout. For the mid-2000s regime shift, $p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0) = 6.9\%$, which represents the effect of factors other than fishing pressure or ocean warming (e.g., eutrophication). This low baseline probability indicates that other factors played a minor role in causing the collapse.

In the "no fishing" scenario WBC is affected by elevated SST due to climate change, but there is zero fishing pressure on the stock. The simulation results are very similar to the baseline scenario, which indicates that ocean warming by itself did not have an important impact on the stock. In fact, for the mid-2000s shift, $p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}) = 6.3\%$ and is thus lower than without climate change. While ecologically counterintuitive, this result can be explained by the fact that increasing SST pushes the stock into a domain of increased stability in our calibrated SCM, irrespective of the level of fishing pressure.

In the "no climate change" scenario the observed fishing pressure on the stock takes place in a counterfactual climate without anthropogenic forcing. This changes the size and stability properties of the stock considerably relative to the baseline scenario. As a result, $p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}_0) = 62.8\%$ for the mid-2000s shift, which indicates that fishing pressure had a large impact on the probability of a regime shift, but was not by itself capable of causing the collapse with certainty.

Causal attribution

In the final step of our causal attribution procedure, we feed the probabilistic information obtained in the counterfactual simulations into an attribution scheme.

In a preliminary step, we attribute the collapse of the WBC stock to the combined effect of overfishing and climate change. Specifically, we compute the fraction of the risk of collapse that is attributable to the combined effect of both factors analogous to climate science (Allen, 2003; Stott et al., 2004; Otto, 2017). It is given by the relative increase in the likelihood of collapse in the presence of fishing pressure and climate change compared to the baseline counterfactual scenario in which both factors are absent (see Methods). Here, this amounts to the increase the probability of regime shift from $p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0) = 6.9\%$ in the absence of both factors to $p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}) = 100\%$ when both factors are present. We find that the extent to which the combined effect of overfishing and climate change has caused the collapse was $1 - p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0)/p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}) = (1 - 0.069/1) \cdot 100\% = 93.7\%$.

To divide the combined effect of both factors into their respective contributions we use the attribution scheme of Mittelstaedt & Baumgärtner (2023). It is based on the game-theoretic Shapley value (Shapley, 1953) which has been developed to determine the contributions of individual firms to the profit earned by a cartel. This scheme considers the relative incremental change in the probability of regime shift Δp_{rs} in the presence of a factor compared to its absence (see Methods). Since this probability change may also depend on the presence or absence of other factors, the scheme considers all potential combinations of presence or absence of all factors. Here, this encompasses the three counterfactual scenarios "baseline", "no fishing" and "no climate change". The respective contribution R of a factor is given by the incremental change due to this factor, averaged over all potential combinations, divided by the factual probability $p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C})$.

The incremental change due to fishing pressure compared to the baseline scenario is $\Delta p_{\mathrm{rs}}(\mathrm{F}) = 55.9\%$. That is, unsustainably high fishing pressure made a regime shift considerably more likely, but was not by itself capable of triggering the shift with certainty. The additional incremental change due to climate change in the presence of fishing is $\Delta p_{\mathrm{rs}}(\mathrm{C}) = 37.2\%$, resulting in a regime shift probability of $p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}) = 100\%$ in the presence of both factors. That is, the additional effect of climate change was to modify the stability properties of the stock such that the shift to the low-productivity regime became inevitable.

Likewise, the incremental change due to climate change compared to the baseline scenario is $\Delta p_{\mathrm{rs}}(\mathrm{C}) = -0.6\%$. While ocean warming changes the stability properties of the stock, this does not by itself increase the likelihood of a regime shift. The additional incremental change due to fishing pressure in the presence of climate change is $\Delta p_{\mathrm{rs}}(\mathrm{F}) = 93.7\%$. The additional impact of fishing pressure is necessary to trigger the regime shift with certainty, which highlights the non-linear effect of the two interacting factors on the likelihood of the collapse. Consequently, we find that the extent to which overfishing has caused the mid-2000s collapse of the WBC stock was 74.8% and climate change 18.3% (Table 3).

Table 3: Causal attribution of the mid-2000s regime shift

Combination	$p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0)$	$\Delta p_{\mathrm{rs}}(\mathrm{F})$	$\Delta p_{\mathrm{rs}}(\mathrm{C})$	$\sum$
F, C_0	6.9%	55.9%	37.2%	100%
F_0, C	6.9%	93.7%	– 0.6%	100%
$R =$	6.9%	**74.8%**	**18.3%**	100%

We also analyze the role of fishing pressure and ocean warming in causing an earlier collapse of the WBC stock in the mid-1980s. We find that the extent to which fishing pressure has caused this collapse was 83.9%, climate change 0%, and other factors 16.1% (Supplementary Table C.3). The model predicts that a regime shift in the mid-1980s was relatively unlikely ($p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}) = 33.3\%$) to occur under the then prevailing conditions of moderate fishing pressure and low SST. The effect of unsustainably high fishing pressure made the regime shift over 6 times more likely compared to the baseline scenario ($p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0) = 4.1\%$).

4.3 Discussion

Our quantitative analysis confirms the qualitative results of previous research on the role of overfishing and climate change in causing the collapse of the WBC fishery (Möllmann et al., 2021; Froese et al., 2022). We have shown that unsustainably high fishing pressure was the main driver of the shift to a low-productivity regime, but was not the sole cause of the collapse. In particular, the contribution of overfishing to causing the collapse was more than four times larger than the contribution of climate change. Nevertheless, our results highlight that progressing climate change altered the stability properties of the stock and was necessary for the high-productivity equilibrium to no longer exist. In particular, warming of the Western Baltic Sea critically altered marine environmental conditions, which had negative consequences for the reproductive success of cod (MacKenzie et al., 2007; Voss et al., 2019). As catch quotas for WBC were consistently set higher than scientifically advised and biologically sustainable (Möllmann et al., 2021), our results indicate a failure of fisheries management to adapt to changing climatic conditions.

Uncertainty propagation

Two sources of uncertainty limit the degree of confidence in our quantitative results.

First, the fisheries data ($\mathrm{SSB}, \mathrm{F}, \mathrm{R}$) used to calibrate the SCM are not empirical measurements, but the output of a statistical stock assessment model (Nielsen & Berg, 2014; Aeberhard et al., 2018) that is based on survey data and reported landings. These models are conditional on assumptions such as constant natural mortality and catchability over time (Ottersen et al., 2013) and their output is subject to considerable uncertainty. While future analyses should be based on data sampled in the field, no better data were available at the time this analysis was conducted.

Second, like with every model, the output of the calibrated SCM is subject to uncertainty. Despite validating the model as suggested in the literature (Grasman et al., 2010) and finding the SCM to be superior to alternative linear or logistic models (Supplementary Table C.2), the estimated coefficients of the model are subject to a statistical error. Hence, variations in the values of both input data and model coefficients could lead to potentially large differences in the predicted stability properties of WBC and the likelihood of collapse, which form the basis of our attribution procedure.

To quantify the effect of uncertainty on the confidence in our results, we repeat the analysis 1,000 times, each time drawing randomly from the sampling distribution of fisheries data and model coefficients (see Methods). This twofold bootstrapping procedure compounds the uncertainty surrounding each of the 3 input variables and 6 model coefficients, which results in a wide distribution of the respective contributions of fishing pressure and climate change to the collapse. For instance, the 90% confidence levels for each factor's respective contribution to causing the mid-2000s regime shift are $\left[20.4\%, 93.4\%\right]$ for fishing pressure, and $\left[2.4\%, 85.6\%\right]$ for climate change. Since the bootstrapped distribution of the respective contributions of both factors also contains model runs in which the timing of the regime shift differs from the main run, these confidence intervals likely underestimate the precision of the results.

Potential other causes

We quantify the respective contributions of overfishing and climate change to causing the collapse of the WBC stock, because these two factors have been identified as the most important ones (Möllmann et al., 2021) and because their respective contributions have been the subject of a long-standing debate. In doing so, we do not explicitly consider other factors that might also have played a role in causing the collapse, e.g. eutrophication or acidification. These factors are still present in the baseline scenario and are thus implicitly included in the baseline probability of regime shift $p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0) = 6.9\%$. This suggests that these factors played only a minor role for the regime shift, despite the big concern about eutrophication of the Baltic Sea that has been expressed already in the Club of Rome report (Meadows et al., 1972).

In principle, the number of factors to which the collapse can be attributed is not limited in our approach. Both the SCM and the attribution scheme of Mittelstaedt & Baumgärtner (2023) can be used with any number of factors. However, increasing the number of factors would place even higher requirements on the quality of model and data. Therefore, including more factors would most likely increase the uncertainty of the attribution.

Counterfactual reference scenarios

We assess the difference that overfishing and climate change made to the likelihood of collapse relative to counterfactual scenarios in which only one or both factors are absent. In general, the choice of reference scenario can have a significant effect on the results and depends on the attribution question (Otto, 2017). To measure the contribution of climate change, a counterfactual climate without anthropogenic forcing is routinely used in climate attribution science (Allen, 2003; Stott et al., 2004). Here, we remove the effect of ocean warming on SST, which is the natural choice of reference scenario. For fishing pressure, it is not obvious what the relevant reference scenario is. For instance, when one aims to attribute the collapse to *over*fishing one might as well consider fishing in exceedance of levels that are deemed sustainable.

To assess how the choice of reference scenario affects the results, we repeat the analysis with two alternative counterfactual reference levels of fishing mortality. The first is $\mathrm{F}_{\mathrm{MSY}}$, which is the current management goal for WBC and refers to the fishing mortality expected to lead to maximum sustainable yield (MSY). Fishing at the MSY level instead of zero fishing in the counterfactual scenarios reduces the contribution of fishing to causing the mid-2000s regime shift to 71.2% (Supplementary Table C.4). The second alternative reference level is F_{pa}, which refers to the "precautionary" fishing mortality that keeps SSB above $\mathrm{B}_{\mathrm{lim}}$ (a biomass reference point below which the reproductive capacity of the stock is impaired) with 95% probability. Using this higher fishing mortality (see Figure 14c and 14d) as the reference level in the counterfactual scenarios further decreases the contribution of fishing to 62.3% (Supplementary Table C.5). Besides these two commonly used reference levels, other (temporally variable) fishing reference levels could be used.

Further attribution questions

We measure the degree to which overfishing and climate change have caused the collapse of WBC in terms of their respective contribution to the likelihood of a shift to a low-productivity regime. This probabilistic approach is well-suited to cases such as WBC where a discontinuous regime shift has been identified as the relevant mechanism behind the stock collapse (Möllmann et al., 2021). In other cases, where the mechanisms underlying the stock collapse are not clear, it may be preferable to attribute the level of SSB directly to overfishing, climate change and other factors (Stecher & Baumgärtner, 2022a). This would also allow answering a wider range of attribution questions, such as what is the role of different factors in reducing the stock size to a particular value, which does not need to be a collapse.

Finally, the concept of causation used here is very general and not limited to abiotic factors in ecosystems, but can be applied more broadly to agents' actions in dynamical systems. For instance, another question related to the collapse of the WBC stock is to what extent the excessive fishing pressure can be attributed to different agents in the fishery, such as fishers, scientific advisers, or fisheries management. This kind of quantitative knowledge is relevant for questions such as who is to blame for the collapse and who should be liable – and to what extent – for the ecological, social and economic damage. In principle, this and other attribution questions can be readily answered with our generic model-based causal attribution procedure.

4.4 Methods

Data

We employ annual data on spawning stock biomass (SSB), recruitment (R) and fishing mortality (F) of WBC from model-based stock assessments by the International Council for the Exploration of the Sea (ICES). Specifically, we use data for cod in ICES subdivisions 22–24 between 1985 and 2021 from the latest official stock assessment (ICES, 2021). We extend this time series backwards until 1970 with data from an earlier assessment (ICES, 2014) to be able to analyze a longer period of time and include relatively high stock sizes in the early 1980s into the calibration. Since the two assessments use different methodologies (the earlier assessment does not consider mixing of WBC and neighboring Eastern Baltic cod), this might induce a bias. However, as stock mixing is considered a more recent phenomenon, we do not expect this bias to affect our results (Möllmann et al., 2021).

For sea surface temperature (SST), we use annual data of mean SST in the third quarter (July to September) in the Western Baltic Sea. For the time period from 1984 to 2019, we use data derived from simulations of the hydrodynamic Kiel Baltic Sea Ice-Ocean Model (Lehmann et al., 2014). To be able to analyze the full time period from 1970 to 2021, we combine this data with lower spatial resolution data for 1970-1983 and 2002-2021. We obtain monthly SST data from the US National Oceanic and Atmospheric Administration (NOAA) Extended Reconstructed Sea Surface Temperature (ERSST) dataset (Huang et al., 2017), from which we compute mean SST in the third quarter. Higher SST in the third quarter are thought to have a negative effect on recruitment by hampering the maturation of spawners (personal communication). For the simulation of counterfactual scenarios, we remove the effect of climate change from the data by detrending the time series of SST. For simplicity, we assume a linear effect of climate change on SST over time.

Software

We conduct the whole analysis using the open-source software *R* (R Core Team, 2021) (version 4.1.3) with the packages *tidyverse* (Wickham et al., 2019) for data handling and graphics, *cusp* (Grasman et al., 2010) for stochastic cusp modeling, and *RConics* (Huber, 2022) for solving cubic equations.

Stochastic cusp model (SCM)

SCM is based on catastrophe theory (Thom, 1975; Zeeman, 1976). The cusp is one of the seven elementary catastrophes identified by Thom (1975) and describes how the dynamic response of a state variable to a control variable changes from linear to discontinuous due to effect of a second interacting control variable (Petraitis, 2013). In its canonical form, the cusp is defined by its dynamic potential $V(Z_t, \alpha, \beta)$, which describes the stability properties of a system with a single state variable Z_t and is given by:

$$V(Z_t, \alpha, \beta) = \frac{1}{4} Z_t^4 - \frac{1}{2} \beta Z_t^2 - \alpha Z_t \ , \tag{42}$$

where α determines the magnitude of Z_t and the location of its equilibria. It is called the 'asymmetry factor', because it determines whether the density function of Z_t is symmetric or skewed. β controls whether the response of Z_t to α is smooth or discontinuous and is known as the 'bifurcation factor', as it determines the number of modes of the density function (Grasman et al., 2010). Together, α and β determine whether the system exhibits a single stable equilibrium or two locally stable equilibria separated by an unstable one, which gives rise to regime shifts and hysteresis. The corresponding differential equation that describes the evolution of the canonical state variable Z_t over time t is obtained by differentiating the negative of (42) with respect to Z_t:

$$\frac{\mathrm{d}Z_t}{\mathrm{d}t} = -\frac{\partial}{\partial Z_t} V(Z_t, \alpha, \beta) = -Z_t^3 + \beta Z_t + \alpha \ . \tag{43}$$

This cubic equation has three real roots if 'Cardan's discriminant' $\delta = 27\alpha - 4\beta^3 < 0$, and one solution if $\delta > 0$. To allow for stochastic diffusion, a standard Wiener process W_t is added to (43), which results in the stochastic differential equation:

$$\mathrm{d}Z_t = \left(-Z_t^3 + \beta_t Z_t + \alpha_t\right) \mathrm{d}t + \sigma \, \mathrm{d}W_t \ , \tag{44}$$

where $\mathrm{d}W_t = W_{t+\mathrm{d}t} - W_t \sim \mathcal{N}\big(0, \mathrm{d}t\big)$ is the infinitesimal increment of the Wiener process. To calibrate this stochastic model to empirical data, the three canonical components Z_t, α and β are first obtained as smooth transformations of the explanatory variables SSB, F/R and SST:

$$Z_t = w_0 + w_1 \cdot \mathrm{SSB}_t \tag{45}$$

$$\alpha_t = \alpha_0 + \alpha_1 \cdot \frac{\mathrm{F}_t}{\mathrm{R}_t} \ . \tag{46}$$

$$\beta_t = \beta_0 + \beta_1 \cdot \mathrm{SST}_t \tag{47}$$

An underlying assumption when modeling the bifurcation factor as a function of SST is that ocean warming modulates the relationship between SSB and F/R from non-linear and discontinuous to linear and continuous. The coefficients $w_0, w_1, \alpha_0, \alpha_1, \beta_0, \beta_1$ are then estimated using maximum likelihood as suggested by Cobb et al. (1983).

Equilibria and probability of regime shift

One can solve for the roots of the cubic equation (44) to obtain the deterministic equilibrium values Z_t^* of the canonical state variable Z_t. If there are three roots, the intermediate unstable equilibrium can be regarded as the threshold value $\widetilde{Z}_t$ beyond which the state variable shifts from the basin of attraction of one locally stable equilibrium to the alternative one. In general, to determine the probability of a stochastic process hitting a known threshold conditional on the process value at an earlier point in time, one integrates over its transition probability density function. However, this function is not known in analytical form for the stochastic cusp model.

As an approximation, the *instantaneous* probability of a regime shift can be calculated by assessing the likelihood of crossing the threshold value $\widetilde{Z}_t$ within the time interval dt due to the combined effect of deterministic drift and stochastic diffusion. If Equation 43 has three roots, the instantaneous probability of regime shift p_{rs} is defined as the conditional probability that a normally distributed random variable with mean Z_t and variance $\sigma^2\,\mathrm{d}t$ is below (above) its threshold value at time $t+1$ given that it was above (below) its threshold value at time t:

$$p_{\mathrm{rs}} = \begin{cases} P(Z_{t+1} < \widetilde{Z}_{t+1}) = P\big(Z_t + \Delta Z\,\mathrm{d}t < \widetilde{Z}_{t+1}\big) & \text{for } Z_t > \widetilde{Z}_t \\ P(Z_{t+1} > \widetilde{Z}_{t+1}) = P\big(Z_t + \Delta Z\,\mathrm{d}t > \widetilde{Z}_{t+1}\big) & \text{for } Z_t < \widetilde{Z}_t \end{cases} \quad \text{if} \;\; \delta_t, \delta_{t+1} < 0\;. \tag{48}$$

If Equation 43 has one root in year $t+1$, the probability of regime shift in year t is either zero or one:

$$p_{\mathrm{rs}} = \begin{cases} 0, & \begin{array}{l} \text{for } Z_t > \widetilde{Z}_t \;\wedge\; Z_{t+1}^* > \widetilde{Z}_t \\ \text{or } \;Z_t < \widetilde{Z}_t \;\wedge\; Z_{t+1}^* < \widetilde{Z}_t \end{array} \\ 1, & \begin{array}{l} \text{for } Z_t > \widetilde{Z}_t \;\wedge\; Z_{t+1}^* < \widetilde{Z}_t \\ \text{or } \;Z_t < \widetilde{Z}_t \;\wedge\; Z_{t+1}^* > \widetilde{Z}_t \end{array} \end{cases} \qquad \text{if} \;\; \delta_t < 0, \delta_{t+1} > 0\,. \tag{49}$$

That is, the probability of shifting to the alternative regime is zero if there is only one and the same single equilibrium in two consecutive years. While the same equilibrium does not necessarily have the same exact numerical value in terms of Z_t, it is the same in qualitative terms – corresponding either to a relatively large stock size above the threshold, or a relatively small one below the threshold. A regime shift is inevitable if the equilibrium in year t ceases to exist in year $t + 1$. The presence or absence of fishing pressure in the counterfactual simulations changes both the predicted value of Z_t as well as the number and location of its equilibria.

Causal attribution

We use the fraction of attributable risk (FAR) to quantify the degree to which the combined effect of overfishing and climate change has caused the collapse of WBC. This measure was originally developed in epidemiology to attribute a disease (e.g., lung cancer) to a prognostic factor (e.g., smoking) (Rothman & Greenland, 2005; Gleiss & Schemper, 2019) and is commonly used in climate science to attribute extreme weather events to climate change (Allen, 2003; Stott et al., 2004; Otto, 2017). In general, the FAR measures the increase in the likelihood of an event due to the presence of a factor relative to its absence. This is in line with counterfactual theories on causation (e.g. Lewis, 1973) and captures how necessary the factor was for the event (Pearl, 2009b). It may be expressed as 1 – probability of the event in the absence of this factor/probability of the event when the factor is present. For attributing the collapse of WBC to the combined effect of overfishing and climate change, it is given by:

$$\mathrm{FAR}(\mathrm{F}, \mathrm{C}) = \frac{p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}) - p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0)}{p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C})} = 1 - \frac{p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0)}{p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C})} \ . \tag{50}$$

To divide the combined effect of both factors on the likelihood of collapse into their respective contributions R, we employ and adapt the attribution scheme of Mittelstaedt & Baumgärtner (2023). In their original setting, there are multiple agents that deliberately and simultaneously take an action that modifies the probability of a regime shift. Here, we treat the factual levels of fishing pressure and ocean warming between 1970 and 2021 as pseudo actions that do not involve agency. Further, we take a probability-reducing effect of climate change into account. While the properties of this attribution scheme do not hold for probability-reducing actions in general, this is not an issue here, because $p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}) = 1$. The attribution scheme is given in general terms in Table 4.

Table 4: Attribution scheme of Mittelstaedt & Baumgärtner (2023) adapted to WBC

Combination	$\Delta p_{\mathrm{rs}}(\mathrm{F})$	$\Delta p_{\mathrm{rs}}(\mathrm{C})$
F, C_0	$p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}_0) - p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0)$	$p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}) - p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}_0)$
F_0, C	$p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C}) - p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C})$	$p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}) - p_{\mathrm{rs}}(\mathrm{F}_0, \mathrm{C}_0)$
$R =$	$\frac{1}{2} \frac{\sum \Delta p_{\mathrm{rs}}(\mathrm{F})}{p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C})}$	$\frac{1}{2} \frac{\sum \Delta p_{\mathrm{rs}}(\mathrm{C})}{p_{\mathrm{rs}}(\mathrm{F}, \mathrm{C})}$

Sensitivity analysis

To obtain the sampling distributions of the fisheries input data variables SSB, R, and F, we compute the respective standard deviation from the reported 95% confidence intervals for these variables in the ICES data. For the model coefficients, we use the standard errors reported in Table C.1 to construct the sampling distributions. We then randomly draw each of the coefficients of the SCM and each of the fisheries data variables 1,000 times from their respective sampling distributions. For each random draw, we simulate the counterfactual scenarios, compute the stability properties of the stock as well as the probability of regime shift, and measure the respective contributions of both factors to the likelihood of collapse as described above. The resulting distribution of the respective contributions of both factors allows us to construct confidence intervals at any confidence level by computing the corresponding quantiles of this distribution.

Acknowledgments

We are grateful to Eyal Frank, Daniele Rinaldo, Rudi Voss, and participants of the IUSWG seminar at the University of Freiburg for discussion and comments, and to the German Federal Ministry of Education and Research (BMBF) for financial support under grant no. 01LC1826C.

Appendix C Supplementary Information

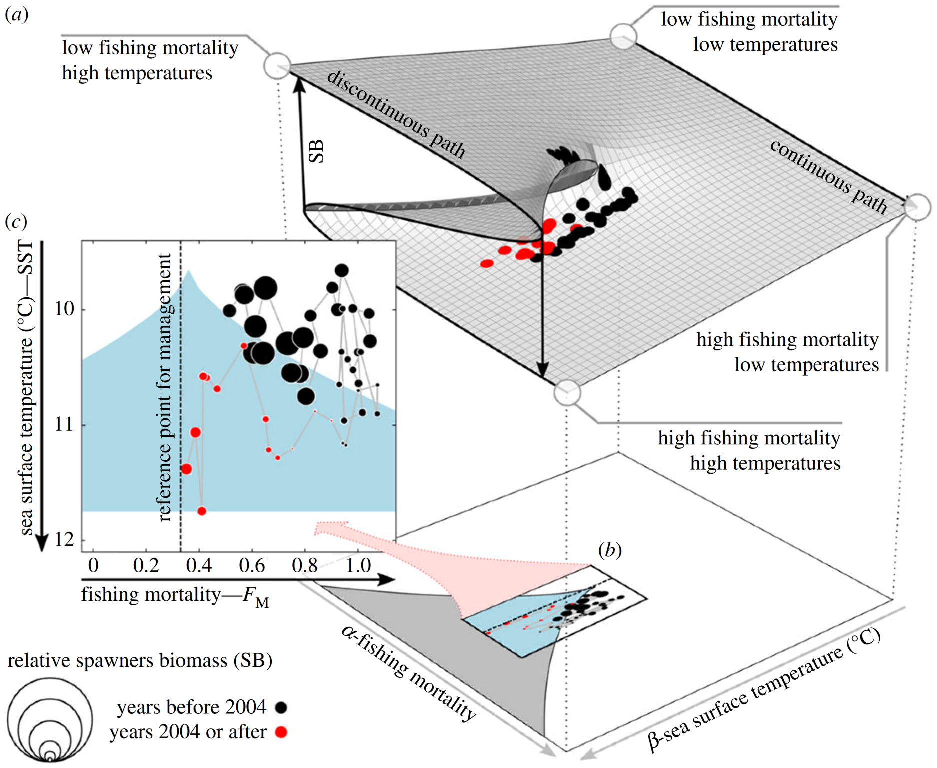


Figure C.1: Two- and three-dimensional representations of the stochastic cusp model. Reprinted from Sguotti et al. (2019) (data for illustrative purposes).

C.1 Model calibration and validation

Table C.1: Estimates of coefficients of stochastic cusp model

Coef.	Estimate	Std. Er.	z	$Pr(>\|z\|)$
α_0	0.478*	0.188	2.545	0.01093
α_1	–15.769**	5.033	–3.133	0.00173
β_0	5.191	3.993	1.300	0.19354
β_1	–0.210	0.231	–0.908	0.36409
w_0	–2.724***	0.242	–11.261	< 0.0001
w_1	0.092**	0.007	13.074	< 0.0001

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.2: Goodness of fit metrics for cusp model and alternative linear and logistic models

	R^2	logLik	N_{par}	AIC	AICc	BIC
Linear model	0.26	–201.26	4	410.51	411.36	418.32
Logistic model	0.59	–185.58	5	381.17	382.47	390.92
Cusp model	0.73	–54.78	6	121.56	123.43	133.27

For the cusp model, R^2 is given by Cobb's pseudo-R^2, which may be negative.

Lower values of AIC, AICc, and BIC indicate a better fit to the data.

C.2 Full results for other collapse and reference scenarios

Table C.3: Causal attribution of the mid-1980s regime shift

Combination	$p_{\text{rs}}(\text{F}_0, \text{C}_0)$	$\Delta p_{\text{rs}}(\text{F})$	$\Delta p_{\text{rs}}(\text{C})$	$\sum$
F, C_0	5.4%	27.9%	0%	100%
F_0, C	5.4%	28.0%	0%	100%
$R =$	16.1%	**83.9%**	**0%**	100%

Table C.4: Causal attribution of the mid-2000s regime shift for F=F$_{\text{MSY}}$ in baseline scenario

Combination	$p_{\text{rs}}(\text{F}_{\text{MSY}}, \text{C}_0)$	$\Delta p_{\text{rs}}(\text{F})$	$\Delta p_{\text{rs}}(\text{C})$	$\sum$
F, C_0	10.3%	52.5%	37.2%	100%
$\text{F}_{\text{MSY}}, \text{C}$	10.3%	89.9%	-0.2%	100%
$R =$	10.3%	**71.2%**	**18.5%**	100%

Table C.5: Causal attribution of the mid-2000s regime shift for F=F$_{\text{pa}}$ in baseline scenario

Combination	$p_{\text{rs}}(\text{F}_{\text{pa}}, \text{C}_0)$	$\Delta p_{\text{rs}}(\text{F})$	$\Delta p_{\text{rs}}(\text{C})$	$\sum$
F, C_0	18.5%	44.3%	37.2%	100%
$\text{F}_{\text{pa}}, \text{C}$	18.5%	80.3%	1.2%	100%
$R =$	18.5%	**62.3%**	**19.2%**	100%

F_{pa} is defined as the level of fishing mortality that leads SSB to the biological reference point B_{pa}, above which the stock is likely to have full reproductive capacity.

References

Aeberhard, W. H., Mills Flemming, J., & Nielsen, A. (2018). Review of state-space models for fisheries science. *Annual Review of Statistics and its Application*, 5, 215–235.

Alberini, A. & Austin, D. (2002). Accidents waiting to happen: liability policy and toxic pollution releases. *Review of Economics and Statistics*, 84(4), 729–741.

Alexander, L. & Moore, M. (2021). Deontological Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2021 edition. https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/.

Allee, W. C., Park, O., Emerson, A. E., Park, T., & Schmidt, K. P. (1949). *Principles of Animal Ecology*. WB Saunders.

Allen, M. (2003). Liability for climate change. *Nature*, 421(6926), 891–892.

Allen, M., Pall, P., Stone, D., & Stott, P. (2007). Scientific challenges in the attribution of harm to human influence on climate. *University of Pennsylvania Law Review*, 155(6), 1353–1400.

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Barnosky, A. D., Hadly, E. A., Bascompte, J., Berlow, E. L., Brown, J. H., Fortelius, M., Getz, W. M., Harte, J., Hastings, A., Marquet, P. A., Martinez, N. D., Mooers, A., Roopnarine, P., Vermeij, G., Williams, J. W., Gillespie, R., Kitzes, J., Marshall1, C., Matzke1, N., Mindell, D. P., Revilla, E., & Smith, A. B. (2012). Approaching a state shift in Earth's biosphere. *Nature*, 486(7401), 52–58.

Bastianoni, S., Pulselli, F. M., & Tiezzi, E. (2004). The problem of assigning responsibility for greenhouse gas emissions. *Ecological Economics*, 49(3), 253–257.

Baumgärtner, S. & Quaas, M. F. (2009). Ecological-economic viability as a criterion of strong sustainability under uncertainty. *Ecological Economics*, 68(7), 2008–2020.

Baumgärtner, S. (2020). Responsibility for regime shifts in managed ecosystems. SSRN Discussion paper 3752006, http://ssrn.com/abstract=3752006.

Baumgärtner, S., Petersen, T., & Schiller, J. (2018). The concept of responsibility: norms, actions and their consequences. SSRN Discussion paper 3157667, https://ssrn.com/abstract=3157667.

Baumgärtner, S. & Quaas, M. (2021). Liability according to actors' responsibility. Presentation at the annual meeting of the German Association of Environmental and Resource Economists (AURÖ), 16–17 Sep 2021.

Beaugrand, G., Balembois, A., Kléparski, L., & Kirby, R. R. (2022). Addressing the dichotomy of fishing and climate in fishery management with the fishclim model. *Communications Biology*, 5(1), 1–13.

Beisner, B. E., Haydon, D. T., & Cuddington, K. (2003). Alternative stable states in ecology. *Frontiers in Ecology and the Environment*, 1(7), 376–382.

Bellman, R. (1966). Dynamic programming. *Science*, 153(3731), 34–37.

Béné, C. & Doyen, L. (2018). From resistance to transformation: a generic metric of resilience through viability. *Earth's Future*, 6(7), 979–996.

Biggs, R., Carpenter, S. R., & Brock, W. A. (2009). Turning back from the brink: detecting an impending regime shift in time to avert it. *Proceedings of the National Academy of Sciences*, 106(3), 826–831.

Blenckner, T., Möllmann, C., Stewart Lowndes, J., Griffiths, J. R., Campbell, E., De Cervo, A., Belgrano, A., Boström, C., Fleming, V., Frazier, M., Neuenfeldt, S., Niiranen, S., Nilsson, A., Ojaveer, H., Olsson, J., Palmlöv, C. S., Quaas, M., Rickels, W., Sobek, A., Viitasalo, M., Wikström, S. A., & Halpern, B. S. (2021). The Baltic Health Index (BHI): Assessing the social–ecological status of the Baltic Sea. *People and Nature*, 3(2), 359–375.

Boomhower, J. (2019). Drilling like there's no tomorrow: bankruptcy, insurance, and environmental risk. *American Economic Review*, 109(2), 391–426.

Bovens, M. (1998). *The Quest for Responsibility: Accountability and Citizenship in Complex Organisations*. Cambridge University Press.

Braham, M. & van Hees, M. (2009). Degrees of causation. *Erkenntnis*, 71(3), 323–344.

Brand, F. S. & Jax, K. (2007). Focusing the meaning(s) of resilience: Resilience as a descriptive concept and a boundary object. *Ecology and Society*, 12(1), 23.

Brander, K. M. (2018). Climate change not to blame for cod population decline. *Nature Sustainability*, 1(6), 262–264.

Bunge, M. A. (1959). *Causality. The Place of the Causal Principle in Modern Science.* Harvard University Press.

Carpenter, S. R. (2003). *Regime Shifts in Lake Ecosystems: Pattern and Variation*, volume 15 of *Excellence in Ecology.* Oldendorf/Luhe: Ecology Institute.

Casey, J. M. & Myers, R. A. (1998). Near extinction of a large, widely distributed fish. *Science*, 281(5377), 690–692.

Chiarella, C., He, X.-Z., & Sklibosios Nikitopoulos, C. (2015). *Derivative Security Pricing. Techniques, Methods and Applications*, volume 21 of *Dynamic Modeling and Econometrics in Economics and Finance.* Springer.

Chockler, H. & Halpern, J. Y. (2004). Responsibility and blame: a structural-model approach. *Journal of Artificial Intelligence Research*, 22, 93–115.

Clark, C. W. (1990). *Mathematical Bioeconomics: The Optimal Management of Renewable Resources.* Wiley, 2$^{\text{nd}}$ edition.

Cobb, L., Koppstein, P., & Chen, N. H. (1983). Estimation and moment recursion relations for multimodal distributions of the exponential family. *Journal of the American Statistical Association*, 78(381), 124–130.

Cobb, L. & Watson, B. (1980). Statistical catastrophe theory: An overview. *Mathematical Modelling*, 1(4), 311–317.

Contamin, R. & Ellison, A. M. (2009). Indicators of regime shifts in ecological systems: What do we need to know and when do we need to know it? *Ecological Applications*, 19(3), 799–816.

Cook, R., Sinclair, A., & Stefansson, G. (1997). Potential collapse of North Sea cod stocks. *Nature*, 385(6616), 521–522.

Cunningham, S. (2021). *Causal Inference: The Mixtape.* Yale University Press.

Dakos, V., Matthews, B., Hendry, A. P., Levine, J., Loeuille, N., Norberg, J., Nosil, P., Scheffer, M., & De Meester, L. (2019). Ecosystem tipping points in an evolving world. *Nature Ecology & Evolution*, 3(3), 355–362.

Daly, E. & Porporato, A. (2006). Probabilistic dynamics of some jump-diffusion systems. *Physical Review E*, 73(2), 026108.

Das, S. R. (2002). The surprise element: jumps in interest rates. *Journal of Econometrics*, 106(1), 27–65.

De Lara, M., Martinet, V., & Doyen, L. (2015). Satisficing versus optimality: criteria for sustainability. *Bulletin of Mathematical Biology*, 77(2), 281–297.

DeAngelis, D. L., Post, W. M., & Travis, C. C. (1986). *Positive Feedback in Natural Systems*, volume 15 of *Biomathematics*. Springer.

deYoung, B., Barange, M., Beaugrand, G., Harris, R., Perry, R. I., Scheffer, M., & Werner, F. (2008). Regime shifts in marine ecosystems: detection, prediction and management. *Trends in Ecology & Evolution*, 23(7), 402–409.

Döring, R., Berkenhagen, J., Hentsch, S., & Kraus, G. (2020). Small-scale fisheries in Germany: a disappearing profession? In *Small-scale Fisheries in Europe: Status, Resilience and Governance* (pp. 483–502).

Doyen, L., Armstrong, C., Baumgärtner, S., Béné, C., Blanchard, F., Cisse, A. A., Cooper, R., Dutra, L., Eide, A., Freitas, D., Gourguet, S., Gusmaoh, P., Hardy, P.-Y., Jarre, R., Little, L., Macher, C., Quaas, M., Regnier, E., Sanz, N., & Thébaud, O. (2019). From no whinge scenarios to viability tree. *Ecological Economics*, 163, 183–188.

Doyen, L. & De Lara, M. (2010). Stochastic viability and dynamic programming. *Systems & Control Letters*, 59(10), 629–634.

Duff, R. (2018). Responsibility. In *Routledge Encyclopedia of Philosophy*. https://www.rep.routledge.com/articles/thematic/responsibility/.

Dutheil, C., Meier, H., Gröger, M., & Börgel, F. (2021). Understanding past and future sea surface temperature trends in the Baltic Sea. *Climate Dynamics*, (pp. 1–19).

D'Odorico, P., Laio, F., & Ridolfi, L. (2006). A probabilistic analysis of fire-induced tree-grass coexistence in savannas. *The American Naturalist*, 167(3), E79–E87.

Epstein, R. A. (1973). A theory of strict liability. *The Journal of Legal Studies*, 2(1), 151–204.

Ferng, J.-J. (2003). Allocating the responsibility of co$_2$ over-emissions from the perspectives of benefit principle and ecological deficit. *Ecological Economics*, 46(1), 121–141.

Ferraro, P. J., Sanchirico, J. N., & Smith, M. D. (2019). Causal inference in coupled human and natural systems. *Proceedings of the National Academy of Sciences*, 116(12), 5311–5318.

Folke, C. (2006). Resilience: The emergence of a perspective for social–ecological systems analyses. *Global environmental change*, 16(3), 253–267.

Folke, C., Carpenter, S., Walker, B., Scheffer, M., Elmqvist, T., Gunderson, L., & Holling, C. S. (2004). Regime shifts, resilience, and biodiversity in ecosystem management. *Annual Review of Ecology, Evolution and Systematics*, 35, 557–581.

Friedman, M. (1961). The lag in effect of monetary policy. *Journal of Political Economy*, 69(5), 447–466.

Froese, R., Papaioannou, E., & Scotti, M. (2022). Climate change or mismanagement? *Environmental Biology of Fishes*. https://doi.org/10.1007/s10641-021-01209-1.

Gaviraghi, B. (2017). *Theoretical and Numerical Analysis of Fokker-Planck Optimal Control Problems for Jump-Diffusion Processes*. PhD thesis, Julius-Maximilians-Universität Würzburg.

Gladwell, M. (2000). *The tipping point: How little things can make a big difference*. Little, Brown.

Gleiss, A. & Schemper, M. (2019). Quantifying degrees of necessity and of sufficiency in cause-effect relationships with dichotomous and survival outcomes. *Statistics in Medicine*, 38(23), 4733–4748.

Good, I. J. (1961). A causal calculus (I). *The British Journal for the Philosophy of Science*, 11(44), 305–318.

Grafton, R. Q., Doyen, L., Béné, C., Borgomeo, E., Brooks, K., Chu, L., Cumming, G. S., Dixon, J., Dovers, S., Garrick, D., Helfgott, A., Jiang, Q., Katic, P., Kompas, T., Little, L. R., Matthews, N., Ringler, C., Squires, D., Steinshamn, S. I., Villasante, S., Wheeler, S., Williams, J., & Wyrwoll, P. R. (2019). Realizing resilience for decision-making. *Nature Sustainability*, 2(10), 907–913.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, (pp. 424–438).

Granger, C. W. & Newbold, P. (1977). *Forecasting Economic Time Series*. Academic Press.

Grasman, R., van der Maas, H. L., & Wagenmakers, E.-J. (2010). Fitting the cusp catastrophe in R: a cusp package primer. *Journal of Statistical Software*, 32, 1–27.

Griffin, J. (1992). The human good and the ambitions of consequentialism. *Social Philosophy and Policy*, 9(2), 118–132.

Gunderson, L. & Holling, C., Eds. (2001). *Panarchy: Understanding Transformations in Human and Natural Systems*. Washington, D.C. USA: Island Press.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11(1), 1–12.

Hannart, A., Pearl, J., Otto, F., Naveau, P., & Ghil, M. (2016). Causal counterfactual theory for the attribution of weather and climate-related events. *Bulletin of the American Meteorological Society*, 97(1), 99–110.

Harrison, G. W. (1979). Stability under environmental stress: resistance, resilience, persistence, and variability. *The American Naturalist*, 113(5), 659–669.

Hart, H. L. A. & Honoré, T. (1959). *Causation in the Law*. Oxford University Press.

Hoegh-Guldberg, O., Jacob, D., Bindi, M., Brown, S., Camilloni, I., Diedhiou, A., Djalante, R., Ebi, K., Engelbrecht, F., Guiot, J., Hijioka, Y., Mehrotra, S., Payne, A., Seneviratne, S., Thomas, A., Warren, R., & Zhou, G. (2018). Impacts of 1.5°C global warming on natural and human systems. In V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. Matthews, Y. Chen, X. Zhou, M. Gomis, E. Lonnoy, T. Maycock, M. Tignor, & T. Waterfield (Eds.), *Global warming of* $1.5°C$ (pp. 175–312). IPCC.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–960.

Holling, C. (1973). Resilience and stability of ecological systems. *Annual Review of Ecological Systems*, 4, 1–23.

Honoré, T. (1995). Necessary and sufficient conditions in tort law. In D. G. Owen (Ed.), *Philosophical Foundations of Tort Law*. Oxford University Press.

Hsieh, C., Glaser, S. M., Lucas, A. J., & Sugihara, G. (2005). Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature*, 435(7040), 336–340.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., & Zhang, H.-M. (2017). NOAA extended reconstructed sea surface temperature (ERSST), version 5. https://www1.ncdc.noaa.gov/pub/data/cmb/ersst/v5/netcdf/.

Huber, E. (2022). RConics: computations on conics. https://cran.r-project.org/web/packages/RConics/index.html.

Hume, D. (1739). *A Treatise of Human Nature.* Oxford University Press.

Hume, D. (1748). *An enquiry concerning human understanding.*

Hutchings, J. A. (2000). Collapse and recovery of marine fishes. *Nature*, 406(6798), 882–885.

International Council for the Exploration of the Seas (ICES) (2014). Report of the Baltic fisheries assessment working group (WGBFAS). ICES CM 2014/ACOM:10.

International Council for the Exploration of the Seas (ICES) (2021). Cod (gadus morhua) in subdivisions 22-24, western Baltic stock (western Baltic Sea). https://doi.org/10.17895/ices.advice.7744.

Jaeger, C. C., Krause, J., Haas, A., Klein, R., & Hasselmann, K. (2008). A method for computing the fraction of attributable risk related to climate damages. *Risk Analysis: An International Journal*, 28(4), 815–823.

Jahn, P., Berg, R. W., Hounsgaard, J., & Ditlevsen, S. (2011). Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *Journal of computational neuroscience*, 31(3), 563–579.

Jiao, Y. (2009). Regime shift in marine ecosystems and implications for fisheries management, a review. *Reviews in Fish Biology and Fisheries*, 19(2), 177–191.

Jordà, Ò., Singh, S. R., & Taylor, A. M. (2022). Longer-run economic consequences of pandemics. *The Review of Economics and Statistics*, 104(1), 166–175.

Keynes, J. M. (1921). *A treatise on probability.* Macmillan & Co.

Kinzig, A. P., Ryan, P., Etienne, M., Allison, H., Elmqvist, T., & Walker, B. H. (2006). Resilience and regime shifts: Assessing cascading effects. *Ecology and Society*, 11(1), 20.

Klein, M. (2005). Responsibility. In T. Honderich (Ed.), *The Oxford companion to philosophy.*

Knight, F. H. (1921). *Risk, uncertainty and profit.* Houghton Mifflin.

Krysiak, F. C. (2011). Nachhaltigkeit, Risiko und Diskontierung. In M. Held, M. Kubon-Gilke, & R. Sturn (Eds.), *Normative und institutionelle Grundfragen der Ökonomik, Jahrbuch 9 — Institutionen ökologischer Nachhaltigkeit* (pp. 133–156). Metropolis Verlag.

Kushner, H. J. & Dupuis, P. G. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*, volume 24 of *Stochastic Modelling and Applied Probability.* Springer.

Lande, R., Engen, S., & Sæther, B.-E. (2003). *Stochastic Population Dynamics in Ecology and Conservation.* Oxford University Press.

Lehmann, A., Hinrichsen, H.-H., Getzlaff, K., & Myrberg, K. (2014). Quantifying the heterogeneity of hypoxic and anoxic areas in the Baltic Sea by a simplified coupled hydrodynamic-oxygen consumption model approach. *Journal of Marine Systems*, 134, 20–28.

Lenton, T. M., Held, H., Kriegler, E., Hall, J., Lucht, W., Rahmstorf, S., & Schellnhuber, H. (2008). Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences*, 105(6), 1786–1793.

Lenzen, M., Murray, J., Sack, F., & Wiedmann, T. (2007). Shared producer and consumer responsibility — Theory and practice. *Ecological Economics*, 61(1), 27–42.

Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70(17), 556–567.

Lewontin, R. (1969). The meaning of stability. In *Diversity and Stability in Ecological Systems* (pp. 12–24). Upton, NY: Brookhaven National Laboratory.

Lindegren, M. & Brander, K. (2018). Adapting fisheries and their management to climate change: a review of concepts, tools, frameworks, and current progress toward implementation. *Reviews in Fisheries Science & Aquaculture*, 26(3), 400–415.

Lotka, A. J. (1925). *Elements of Physical Biology.* Williams & Wilkins.

Lovejoy, T. E. & Nobre, C. (2018). Amazon tipping point. *Science Advances*, 4(2), eaat2340.

Ludwig, D., Walker, B., & Holling, C. S. (1997). Sustainability, stability, and resilience. *Conservation Ecology*, 1(1).

MacKenzie, B. R., Gislason, H., Möllmann, C., & Köster, F. W. (2007). Impact of 21st century climate change on the Baltic Sea fish community and fisheries. *Global Change Biology*, 13(7), 1348–1367.

Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly*, 2(4), 245–264.

May, R. M. (1977). Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature*, 269, 471–477.

Meadows, D. H., Meadows, D. H., Randers, J., & Behrens III, W. W. (1972). *The Limits to Growth*. Universe Books.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1–2), 125–144.

Mill, J. S. (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, volume 1. John W. Parker, London.

Mitroff, I. I. & Silvers, A. (2013). Probabilistic causality. *Technological Forecasting and Social Change*, 80(8), 1629–1634.

Mittelstaedt, C. & Baumgärtner, S. (2023). Attribution of collective causal responsibility to individual actors in stochastic systems. SSRN Discussion paper 3967091, https://ssrn.com/abstract=3967091.

Möllmann, C., Cormon, X., Funk, S., Otto, S. A., Schmidt, J. O., Schwermer, H., Sguotti, C., Voss, R., & Quaas, M. (2021). Tipping point realized in cod fishery. *Scientific Reports*, 11(1), 1–12.

Möllmann, C., Folke, C., Edwards, M., & Conversi, A. (2015). Marine regime shifts around the globe: theory, drivers and impacts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1659), 20130260.

Myers, R. A., Hutchings, J. A., & Barrowman, N. J. (1997). Why do fish stocks collapse? The example of cod in Atlantic Canada. *Ecological applications*, 7(1), 91–106.

Mäler, K.-G., Li, C.-Z., & Destouni, G. (2007). Pricing resilience in a dynamic economy-environment system: A capital-theoretic approach. *Beijer Discussion Paper Series*, 208.

Neumayer, E. (2003). *Weak versus Strong Sustainability: Exploring the Limits of Two Opposing Paradigms*. Edward Elgar Publishing.

Neyman, J. (1923, 1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, 5(4), 465–472. Translated by D.M. Dabrowska and T.P. Speed.

Nielsen, A. & Berg, C. W. (2014). Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research*, 158, 96–101.

Noy-Meir, I. (1975). Stability of grazing systems: an application of predator–prey graphs. *Journal of Ecology*, 63, 459–481.

Ottersen, G., Stige, L. C., Durant, J. M., Chan, K.-S., Rouyer, T. A., Drinkwater, K. F., & Stenseth, N. C. (2013). Temporal shifts in recruitment dynamics of North Atlantic fish stocks: effects of spawning stock and temperature. *Marine Ecology Progress Series*, 480, 205–225.

Otto, F. E. L. (2017). Attribution of weather and climate events. *Annual Review of Environment and Resources*, 42, 627–642.

Oubraham, A. & Zaccour, G. (2018). A survey of applications of viability theory to the sustainable exploitation of renewable resources. *Ecological economics*, 145, 346–367.

Palmer, M. C., Deroba, J. J., Legault, C. M., & Brooks, E. N. (2016). Comment on "slow adaptation in the face of rapid warming leads to collapse of the gulf of maine cod fishery". *Science*, 352(6284), 423–423.

Pearl, J. (2009a). Causal inference in statistics: an overview. *Statistics Surveys*, 3, 96–146.

Pearl, J. (2009b). *Causality*. Cambridge University Press.

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., Nye, J. A., Record, N. R., Scannell, H. A., Scott, J. D., Sherwood, G. D., & Thomas, A. C. (2015). Slow adaptation in the face of rapid warming leads to collapse of the gulf of maine cod fishery. *Science*, 350(6262), 809–812.

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., Nye, J. A., Record, N. R., Scannell, H. A., Scott, J. D., Sherwood, G. D., & Thomas, A. C. (2016). Response to comments on "slow adaptation in the face of rapid warming leads to collapse of the gulf of maine cod fishery". *Science*, 352(6284), 423–423.

Petraitis, P. (2013). *Multiple Stable States in Natural Ecosystems*. Oxford, UK: Oxford University Press.

Pfrommer, T., Goeschl, T., Proelss, A., Carrier, M., Lenhard, J., Martin, H., Niemeier, U., & Schmidt, H. (2019). Establishing causation in climate litigation: admissibility and reliability. *Climatic Change*, 152(1), 67–84.

Pimm, S. L. (1984). The complexity and stability of ecosystems. *Nature*, 307(5949), 321–326.

Pindyck, R. S. (1984). Uncertainty in the theory of renewable resource markets. *The Review of Economic Studies*, 51(2), 289–303.

Pitchford, R. (1995). How liable should a lender be? The case of judgment-proof firms and environmental risk. *American Economic Review*, 85(5), 1171–1186.

Privault, N. (2013). *Stochastic Finance: An Introduction with Market Examples*. Chapman and Hall/CRC.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reichenbach, H. (1956). *The Direction of Time*. University of California Press.

Rinaldi, S. & Scheffer, M. (2000). Geometric analysis of ecological models with slow and fast processes. *Ecosystems*, 3(6), 507–521.

Rodrigues, J., Domingos, T., Giljum, S., & Schneider, F. (2006). Designing an indicator of environmental responsibility. *Ecological Economics*, 59(3), 256–266.

Rosier, S. H. R., Reese, R., Donges, J. F., De Rydt, J., Gudmundsson, G. H., & Winkelmann, R. (2021). The tipping points and early warning indicators for Pine Island Glacier, West Antarctica. *The Cryosphere*, 15(3), 1501–1516.

Rothman, K. J. & Greenland, S. (2005). Basic concepts. In W. Ahrens & I. Pigeot (Eds.), *Handbook of Epidemiology* (pp. 43–88).

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371), 591–593.

Scheffer, M. (2009). *Critical transitions in nature and society*. Princeton University Press.

Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., & Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature*, 413, 591–596.

Scheffer, M. & Carpenter, S. R. (2003). Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in Ecology & Evolution*, 18(12), 648–656.

Schuss, Z. (2010). *Theory and Applications of Stochastic Processes. An Analytical Approach*, volume 170 of *Applied Mathematical Sciences*. Springer.

Sguotti, C., Otto, S. A., Frelat, R., Langbehn, T. J., Ryberg, M. P., Lindegren, M., Durant, J. M., Chr. Stenseth, N., & Möllmann, C. (2019). Catastrophic dynamics limit Atlantic cod recovery. *Proceedings of the Royal Society B*, 286(1898), 20182877.

Shapley, L. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games, Volume II* (pp. 307–318). Princeton University Press.

Shavell, S. (1987). *Economic Analysis of Accident Law*. Harvard University Press.

Shepherd, T. G. (2016). A common framework for approaches to extreme event attribution. *Current Climate Change Reports*, 2(1), 28–38.

Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E., & Knutti, R. (2020). Climate change now detectable from any single day of weather at global scale. *Nature Climate Change*, 10(1), 35–41.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1), 65–94.

Stecher, M. & Baumgärtner, S. (2022a). Quantifying agents' responsibility: a generalized measure of causation in dynamical systems. SSRN Discussion paper 4277765, https://ssrn.com/abstract=4277765.

Stecher, M. & Baumgärtner, S. (2022b). A stylized model of stochastic ecosystems with alternative stable states. *Natural Resource Modeling*, 35(4), e12345.

Steffen, W., Rockström, J., Richardson, K., Lenton, T. M., Folke, C., Liverman, D., Summerhayes, C., Barnosky, A., Cornell, S., Crucifix, M., Donges, J., Fetzer, I., Lade, S. J., Scheffer, M., Winkelmann, R., & Schellnhuber, H. (2018). Trajectories of the earth system in the anthropocene. *Proceedings of the National Academy of Sciences*, 115(33), 8252–8259.

Stern, R. (2004). Does 'ought' imply 'can'? And did Kant think it does? *Utilitas*, 16(1), 42–61.

Stott, P. A., Stone, D. A., & Allen, M. R. (2004). Human contribution to the European heatwave of 2003. *Nature*, 432(7017), 610–614.

Strunz, S. (2012). Is conceptual vagueness an asset? Arguments from philosophy of science applied to the concept of resilience. *Ecological Economics*, 76, 112–118.

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, 338(6106), 496–500.

Swain, D. P., Benoît, H. P., Cox, S. P., & Cadigan, N. G. (2016). Comment on "slow adaptation in the face of rapid warming leads to collapse of the gulf of maine cod fishery". *Science*, 352(6284), 423–423.

Talbert, M. (2022). Moral Responsibility. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Fall 2022 edition. https://plato.stanford.edu/archives/fall2022/entries/moral-responsibility/.

Thom, R. (1975). *Structural Stability and Morphogenesis (translated by DH Fowler)*. W.A. Benjamin, New York.

Uhlenbeck, G. E. & Ornstein, L. S. (1930). On the theory of Brownian Motion. *Physical Review*, 36(5), 823.

Vallentyne, P. (2008). Brute luck and responsibility. *Politics, Philosophy & Economics*, 7(1), 57–80.

van den Bergh, J. C. (2014). Sustainable development in ecological economics. In G. Atkinson, S. Dietz, E. Neumayer, & M. Agarwala (Eds.), *Handbook of Sustainable Development* (pp. 41–54). Edward Elgar Publishing.

Van Inwagen, P. (1978). Ability and responsibility. *The Philosophical Review*, 87(2), 201–224.

van Nes, E. H., Arani, B. M., Staal, A., van der Bolt, B., Flores, B. M., Bathiany, S., & Scheffer, M. (2016). What do you mean, 'tipping point'? *Trends in Ecology & Evolution*, 31(12), 902–904.

Vert-Pre, K. A., Amoroso, R. O., Jensen, O. P., & Hilborn, R. (2013). Frequency and intensity of productivity regime shifts in marine fish stocks. *Proceedings of the National Academy of Sciences*, 110(5), 1779–1784.

Volterra, V. (1926). Fluctuations in the abundance of a species considered mathematically. *Nature*, 118(2972), 558–560.

Voss, R., Quaas, M. F., Stiasny, M. H., Hänsel, M., Pinto, G. A. S. J., Lehmann, A., Reusch, T. B., & Schmidt, J. O. (2019). Ecological-economic sustainability of the Baltic cod fisheries under ocean warming and acidification. *Journal of Environmental Management*, 238, 110–118.

Walker, B., Holling, C., Carpenter, S., & Kinzig, A. (2004). Resilience, adaptability and transformability in social–ecological systems. *Ecology and Society*, 9(2).

Wickham, H. et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557–585.

Zeeman, E. C. (1976). Catastrophe theory. *Scientific American*, 234(4), 65–83.