Ecological Economics 217 (2024) 108086

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Analysis

Quantifying agents' causal responsibility in dynamical systems

Michael Stecher *, Stefan Baumgärtner

Chair of Environmental Economics and Resource Management, University of Freiburg, Germany

ARTICLE INFO

JEL classification:
Q57

Keywords:
Attribution
Causation
Dynamical systems
Responsibility

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.

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[1] 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 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

* Corresponding author.
E-mail addresses: michael.stecher@isi.fraunhofer.de (M. Stecher), stefan.baumgaertner@ere.uni-freiburg.de (S. Baumgärtner).

[1] 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).

https://doi.org/10.1016/j.ecolecon.2023.108086
Received 17 November 2022; Received in revised form 12 December 2023; Accepted 13 December 2023
Available online 19 December 2023

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 and 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)[2] and "excludability".[3] 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 and 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.

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 and Honoré, 1959; Good, 1961; Mackie, 1965; Lewis, 1973; Pearl, 2009). 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 (2009) 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 and 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 (2009) approach does not ascribe causality to agents and their actions. Gleiss and 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 and 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 and 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 and 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 to which a given outcome in a dynamical system is attributed to an agent's action. 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 the realization of 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

[2] 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).

[3] 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).

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 2 we present a simple and general setup of stochastic dynamical systems, which forms the basis of our analysis. In Section 3 we review established philosophical ideas on causation and develop a quantitative measure of causal responsibility. In Section 4 we apply this measure to a number of dynamical systems and different management actions. In Section 5 we highlight the relevance of our concept and its implications for normative responsibility. In the final Section 6 we discuss limitations and conclude.

2. Model and setup

The evolution of the system state[4] $X_t \in [0,\infty)$ over time $t \in [0,\infty)$ is the realization of a stochastic process X_t,[5] which is described by a stochastic differential equation of form

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}Z_t \ , \tag{1}$$

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 Eq. (1) analytically or numerically. Suppose that the solution over the entire time interval $[0,\infty)$ is known. We assume that the stochastic process X_t (Eq. (1)) satisfies the Markov property and admits a stationary probability distribution.

Given the stochastic dynamics (1) of the system state X_t, there exists a stationary probability density function $p(x)$. Hence, the probability that a realization of the process X_t lies in the interval $[\underline{x},\bar{x}] \subseteq [0,\infty)$ at some time t is given by:

$$\mathrm{Prob}\big(X_t \in [\underline{x},\bar{x}]\big) = \int_{\underline{x}}^{\bar{x}} p(x)\,\mathrm{d}x \ . \tag{2}$$

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

$$\mathrm{Prob}\big(X_t \in [\underline{x},\bar{x}] \,\big|\, x_0\big) = \int_{\underline{x}}^{\bar{x}} p(x,t \mid x_0,0)\,\mathrm{d}x =: P_{X_t}(\underline{x},\bar{x}) \ , \tag{3}$$

where the last expression is introduced to simplify notation and denotes the integral of the conditional probability density associated to the process X_t.

Actions

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

$$\mathrm{d}X_t^a = f(X_t, a)\,\mathrm{d}t + g(X_t, a)\,\mathrm{d}Z_t \ . \tag{4}$$

Consequently, the probabilities (2) and (3) 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 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 $p(x,t \mid x_0^a, 0)$. 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 stationary 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 stationary 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:** $\mathbb{E}[X_t^a \mid x_0^a] \neq \mathbb{E}[X_t | x_0]$, $\quad \mathbb{E}[X_t^a] = \mathbb{E}[X_t]$
Rate modifications change the conditional mean, but do not affect its stationary mean. In particular, rate modifications alter the speed and variability of the convergence process towards the stationary 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 stationary 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 4.

[4] 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.

[5] 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.

3. Conceptualizing and measuring causal responsibility

Causal responsibility ascribes the consequences of an action to its perpetrator.[6] 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 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 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.

3.1. 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, 2009, Chap. 7), such as in Section 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 and van Hees, 2009; Pearl, 2009; Mitroff and Silvers, 2013; Gleiss and 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 and 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 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 and Halpern, 2004; Vallentyne, 2008; Braham and 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.[7]

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.

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

[7] 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, 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 and Honoré, 1959; Honoré, 1995).

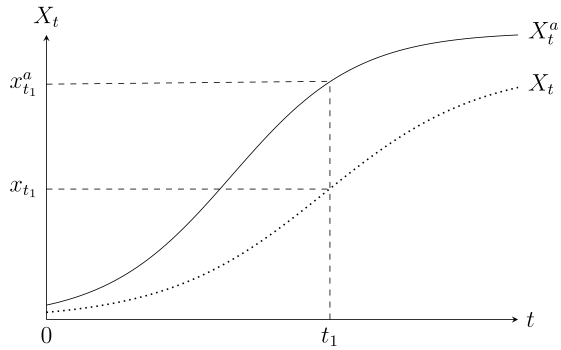

Fig. 1. Intuition of measuring causal responsibility in deterministic systems: actual system state X_t^a (Eq. (4) with $g(X_t, a) = 0$) and counterfactual system state X_t (Eq. (1) with $g(X_t) = 0$) over time.

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, 2009). Alternatively, one may ask prospectively[8] 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.

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.

3.2. Causal responsibility in deterministic systems

Suppose the state of some deterministic dynamical system is $x_{t_1}^a$ 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 $x_{t_1}^a$ 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 3.1, 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 $x_{t_1}^a$ 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 (Fig. 1) is uniquely determined by Eq. (1) with $g(X_t) = 0$ for all X_t.

Definition 1. An agent's causal responsibility for the realized deterministic system state x_t^a 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:

$$R(x_t^a, x_t) = \frac{|\, x_t^a - x_t \,|}{\max\{x_t^a, x_t\}} \tag{5}$$

The numerator of (5) 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 (5) has been introduced from an ex-post perspective, but formally also holds for the ex-ante perspective.

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 (Fig. 2). In addition, the counterfactual process X_t in the absence of action has also infinitely many other potential realizations.

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 (4), the counterfactual system state that would have been realized in the absence of action a can be uniquely determined as follows[9]: 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 Eq. (4). This particular realization Z_t of the stochastic forcing is then used to simulate the counterfactual realization X_t by inserting Z_t into Eq. (1).[10]

Beyond sufficiency, in stochastic systems one also needs to consider how necessary the action was for the realized system state and to

[8] 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).

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

[10] 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.

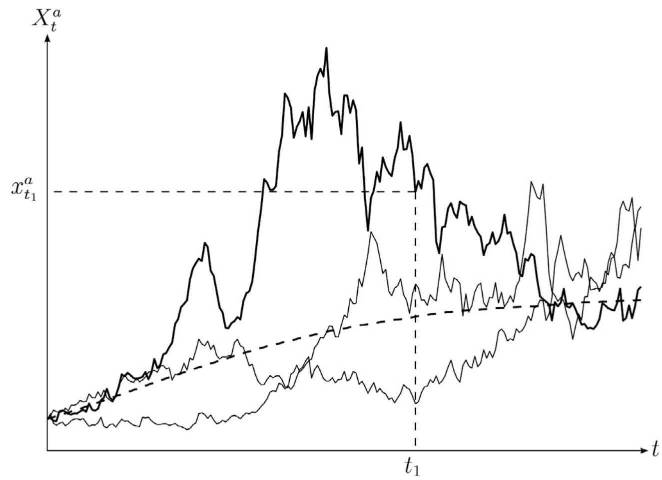

Fig. 2. Epistemological problem in stochastic dynamical systems: three random realizations of the stochastic process X_t^a (described by Eq. (4)) and its expected value (dashed curve). In practice, one only observes a single realization, such as the one drawn in bold, with corresponding system state $x_{t_1}^a$ at time t_1.

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 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'.[11]

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 3.1, 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 2. 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:

$$R\left(x_t^a, x_t\right) = \begin{cases} \dfrac{P_{X_t^a}(\underline{x}, \bar{x}) - P_{X_t}(\underline{x}, \bar{x})}{P_{X_t^a}(\underline{x}, \bar{x})} \cdot \dfrac{|\, x_t^a - x_t \,|}{\max\{x_t^a, x_t\}} & \text{for} \quad P_{X_t^a}(\underline{x}, \bar{x}) > P_{X_t}(\underline{x}, \bar{x}) \\ 0 & \text{for} \quad P_{X_t^a}(\underline{x}, \bar{x}) \le P_{X_t}(\underline{x}, \bar{x}) \end{cases} . \tag{6}$$

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.[12] 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 (6) reduces to the deterministic measure (5) presented in Section 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 3. An agent's ex-ante expected causal responsibility at time t for taking action a is given by:

$$R^e(a) = \mathbb{E}\left[R(x_t^a, x_t)\right] . \tag{7}$$

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

4. Application and results

In this section, we apply the measures (5), (6) and (7) 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 4.3, we present general results and conjectures that follow from the examples presented here and are also informed by our more encompassing analysis.[13]

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:

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = rX_t\left(1 - \frac{X_t}{K}\right) , \tag{8}$$

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. Eq. (8) has a single, stable non-trivial equilibrium at $X_t = K$. In this model, the elementary action types presented in Section 2 correspond to modifying, at time 0, the values of r (rate modification) and K (attractor modification), which affect the 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 Eq. (8).

[11] This is equivalent to the systematic attribution procedure presented by Baumgärtner (2020). In this procedure, a fraction of $\left[P_{X_t^a}(\underline{x}, \bar{x}) - P_{X_t}(\underline{x}, \bar{x})\right] / P_{X_t^a}(\underline{x}, \bar{x})$ 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.

[12] 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).

[13] 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 and Baumgärtner, 2022).

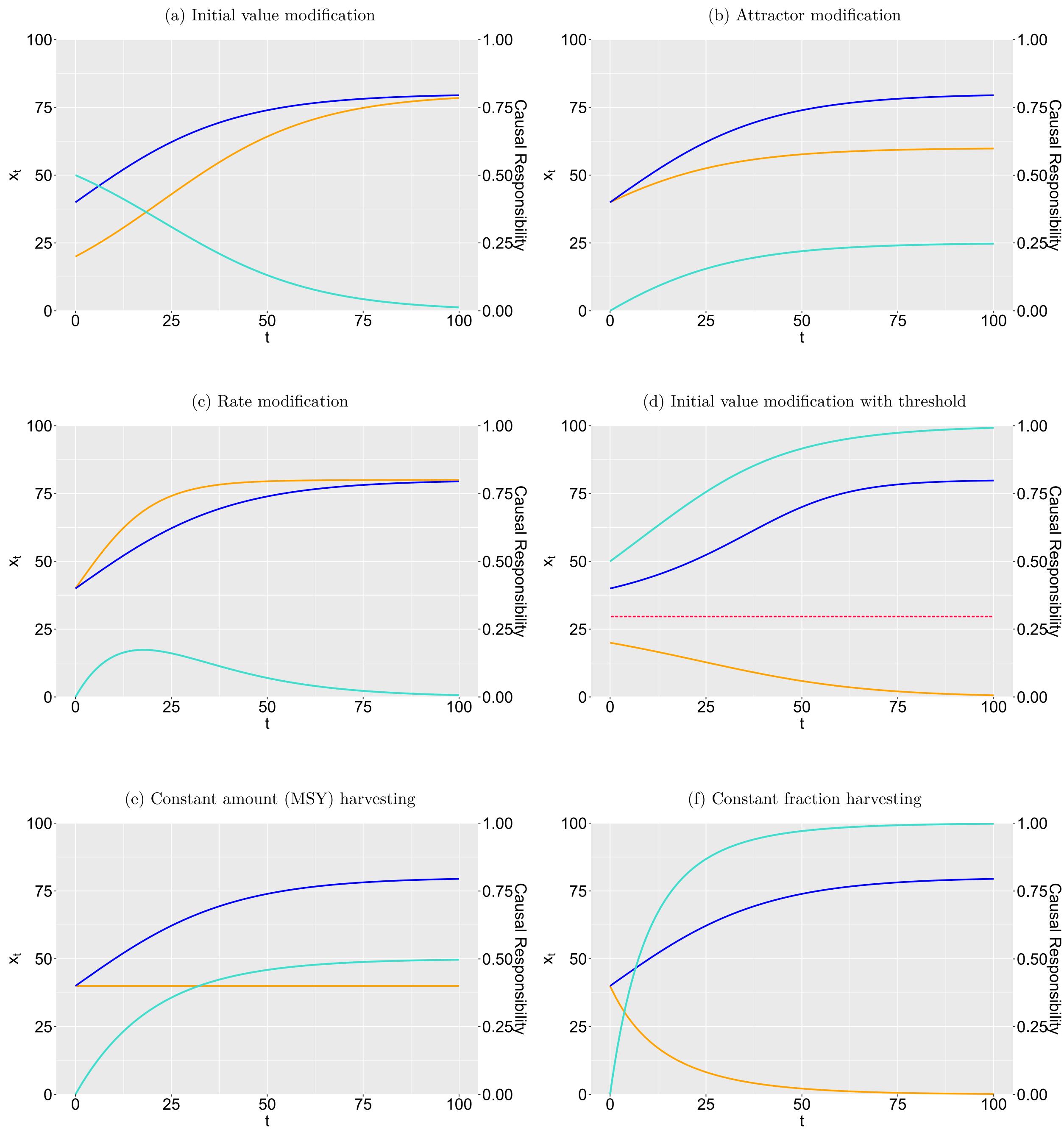

Fig. 3. Evolution of the actual system state X_t^a (solid orange), counterfactual system state X_t (solid blue) and causal responsibility $R(x_t^a, x_t)$ (Eq. (5)) (solid turquoise) over time under deterministic logistic stock dynamics with and without threshold (Eq. (8) for a,b,c; Eq. (9) for d) for 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 (Eq. (5)) for the system state is maximal at time 0 and subsequently decreases over time (Fig. 3(a)). 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 (Fig. 3(b)). 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 (Fig. 3(c)).

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 (Fig. 3(e)). 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.

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 (Fig. 3(f)). 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):

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = rX_t\left(1 - \frac{X_t}{K}\right)\left(\frac{X_t}{V} - 1\right) . \tag{9}$$

The stability properties of Eq. (9) are different from those of Eq. (8): in addition to the stable equilibrium at $X_t = K$, Eq. (9) 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 (Fig. 3(d)).

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:

$$\mathrm{d}X_t = rX_t\left(1 - \frac{X_t}{K}\right)\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t , \tag{10}$$

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 (Eq. (6)) 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 (Fig. A.1). An agent's ex-post causal responsibility (Eq. (6)) for the actually realized system state x_t^a at time t depends on the specific, random realization of the stochastic process described by Eq. (10).

Fig. 4(a) 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 (Eq. (6)) decreases over time.

Fig. 4(b) shows the agent's ex-ante causal responsibility (Eq. (7)) 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 Fig. 3(a), 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 (Fig. 4(a)) is smoothed over by averaging over a large number of realizations due to the law of large numbers.[14] 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.

Further, the agent can modify the system's susceptibility to stochastic shocks by changing the value of σ (Fig. 4(e)). Increasing σ leads to a higher variance of the process X_t^a (Eq. (10)) 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 (Fig. 4(f)). 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.

Fig. 5 depicts the case of a natural resource with economically optimal harvesting that maximizes discounted net surplus for isoelastic

[14] 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.

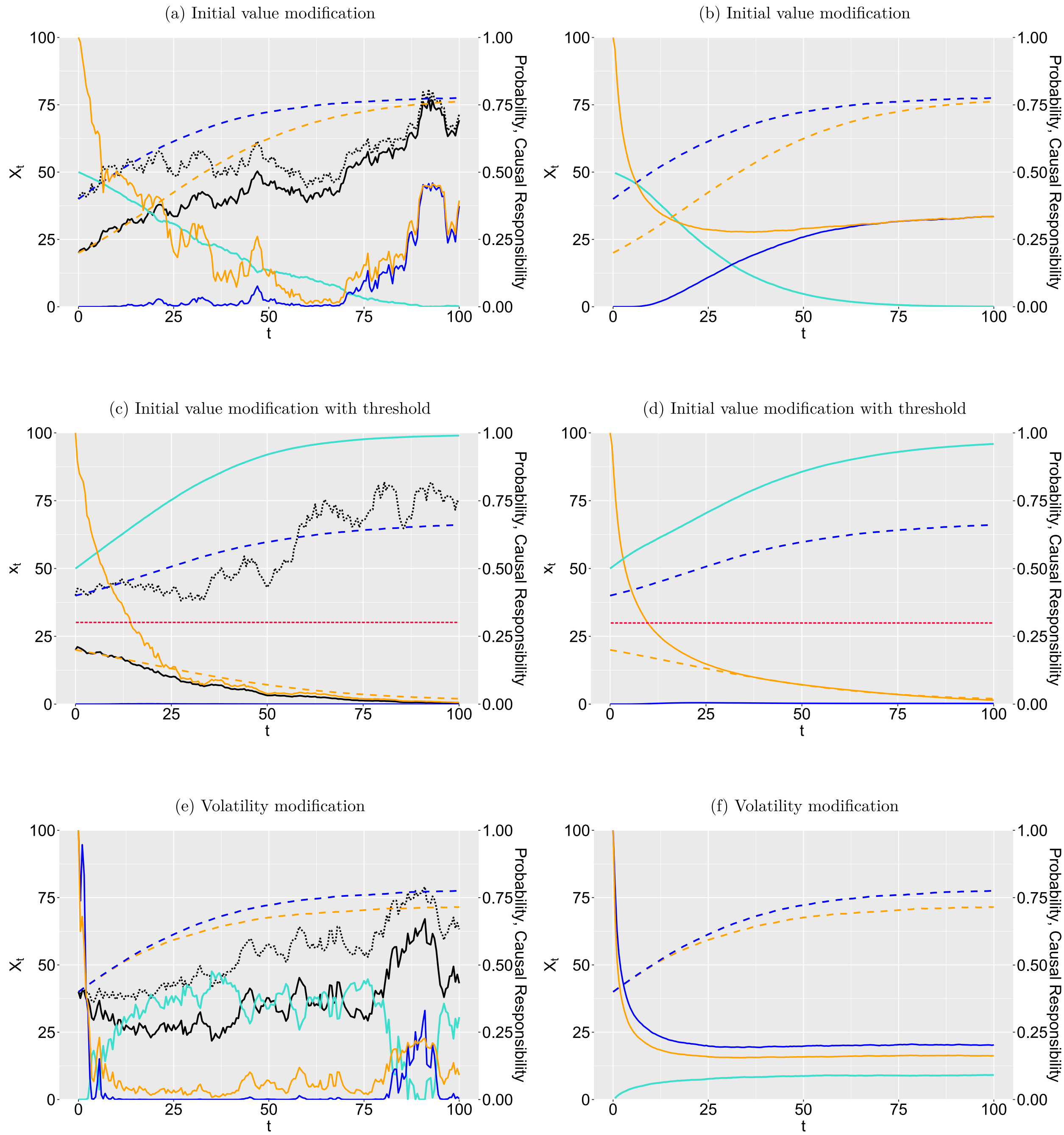

Fig. 4. Realized system state X_t^a (solid black), corresponding expected value $\mathbb{E}[X_t^a]$ (dashed orange) and probability $P_{X_t^a}(\underline{x}, \bar{x})$ (solid orange), counterfactual realization X_t (dotted black), corresponding expected value $\mathbb{E}[X_t]$ (dashed blue) and probability $P_{X_t}(\underline{x}, \bar{x})$ (solid blue), as well as ex-post causal responsibility $R(x_t^a, x_t)$ (Eq. (6), panels a,c,e) and ex-ante causal responsibility $R^e(a)$ (Eq. (7), panels b,d,f) (both solid turquoise) under stochastic logistic stock dynamics with and without thresholds (Eq. (10) for a,b,e,f; Eq. (12) 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

demand and marginal cost functions.[15] 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 (Fig. 5(a)). In general, causal responsibility is relatively high when choosing an optimal control strategy, because the system is exploited to a strong degree (Figs. 5(a) and 5(b)). 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 (Figs. 5(c) and 5(d)).

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

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

$$h^*(X_t) = bX_t \left\{ c + \frac{2b^2 + 2b\left[b^2 + c\left(r + \delta - \sigma^2\right)^2\right]^{\frac{1}{2}}}{\left(r + \delta - \sigma^2\right)^2} \right\}^{-\frac{1}{2}}, \tag{11}$$

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.

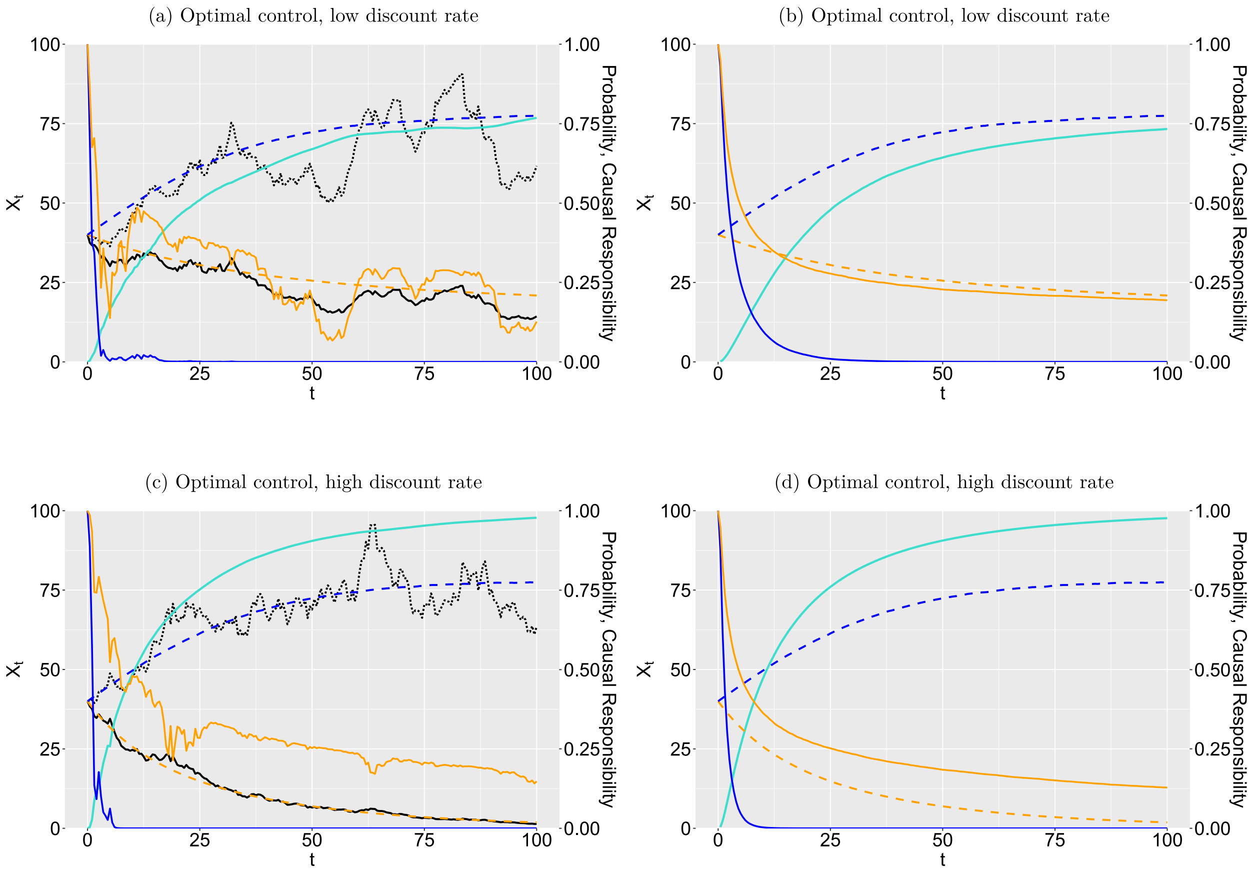

Fig. 5. Realized system state X_t^a (solid black), corresponding expected value $\mathbb{E}[X_t^a]$ (dashed orange) and probability $P_{X_t^a}(\underline{x}, \bar{x})$ (solid orange), counterfactual realization X_t (dotted black), corresponding expected value $\mathbb{E}[X_t]$ (dashed blue) and probability $P_{X_t}(\underline{x}, \bar{x})$ (solid blue), as well as ex-post causal responsibility $R(X_t^a, X_t)$ (Eq. (6), panel a) and ex-ante causal responsibility $R^e(a)$ (Eq. (7), panel b) (both solid turquoise) under stochastic logistic stock dynamics (Eq. (10)) and economically optimal harvesting $h^*(X_t)$ (Eq. (11)). 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\mathrm{d}X_t = rX_t\left(1 - \frac{X_t}{K}\right)\left(\frac{X_t}{V} - 1\right)\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t\ . \tag{12}$$

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 the threshold (Fig. 4(c)).[16] 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.

Similar to the deterministic case (Fig. 3(d)), 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 (Fig. 4(d)). 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.

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 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 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.

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 1 summarizes these results.

[16] 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.

Table 1
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.

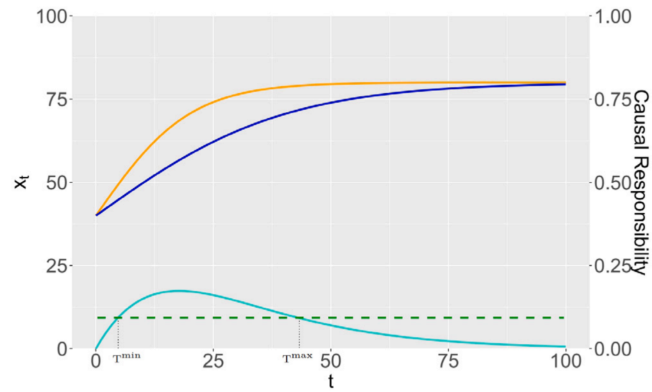

Fig. 6. Actual system state X_t^a (solid orange), counterfactual system state X_t (solid blue) and causal responsibility $R(X_t^a, X_t)$ (Eq. (5)) (solid turquoise) under deterministic logistic stock dynamics (Eq. (8)) with significance threshold $\overline{R}$ (dashed green). Parameter values: $r = 0.05, r^a = 0.1, K = 80, x_0 = 40, \overline{R} = 0.1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 $\mathcal{T}^{\mathrm{sig}}(a)$ during which an action a's causal impact on the system is significant is defined by:

$$\mathcal{T}^{\mathrm{sig}}(a) := \left\{ t \mid R^e\left(a\right) \geq \overline{R} \right\} . \tag{13}$$

For cases of vanishing causal responsibility there exists some $T^{\max}(a) := \sup \mathcal{T}^{\mathrm{sig}}(a)$. After this point in time action a no longer exerts a significant causal influence on the system (Fig. 6).[17] It describes the maximum temporal extent of an action's causal efficacy on the system. Likewise, there exists a minimum time $T^{\min}(a) := \inf \mathcal{T}^{\mathrm{sig}}(a)$ 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.

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.

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.

5.2. Liability

Our concept is relevant for the design of strict[18] 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 establish appropriate incentives for an efficient allocation (Shavell, 1987; Pitchford, 1995; Alberini and 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 and 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.

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

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

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.

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.[19] 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 (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 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 1), 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.

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 multidimensional 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 and Halpern, 2004; Braham and van Hees, 2009; Mittelstaedt and 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.

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

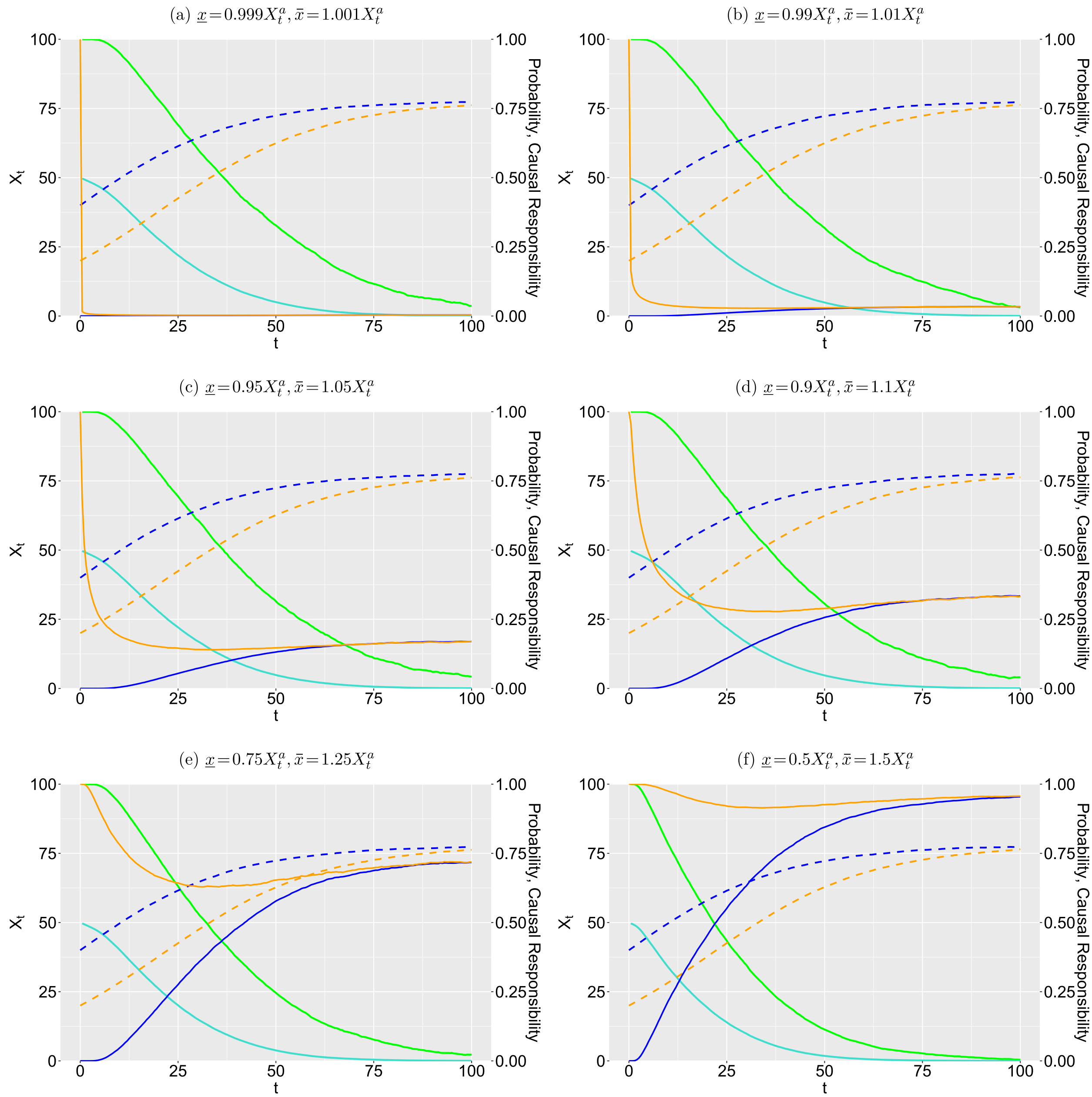

Fig. A.1. Actual expected value $\mathbb{E}[X_t^a]$ and probability $P_{X_t^a}(\underline{x}, \bar{x})$, counterfactual expected value $\mathbb{E}[X_t]$ and probability $P_{X_t}(\underline{x}, \bar{x})$, as well as ex-ante causal responsibility $R^e(a)$ (Eq. (7)) and degree of necessity DN (first factor in Eq. (6)) under stochastic logistic stock dynamics (Eq. (10)). 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.

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.

CRediT authorship contribution statement

Michael Stecher: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Stefan Baumgärtner:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The co-author Stefan Baumgärtner is Co-Editor-in-Chief of the journal Ecological Economics.

Data availability

No data was used for the research described in the article.

Acknowledgments

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

See Fig. A.1.

References

Alberini, A., Austin, D., 2002. Accidents waiting to happen: liability policy and toxic pollution releases. Rev. Econ. Stat. 84 (4), 729–741.

Alexander, L., Moore, M., 2021. Deontological ethics. In: Zalta, E.N. (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.

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

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

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. In: Presentation At the Annual Meeting of the German Association of Environmental and Resource Economists (AURÖ). vol. 2021, pp. 16–17.

Boomhower, J., 2019. Drilling like there's no tomorrow: bankruptcy, insurance, and environmental risk. Amer. Econ. Rev. 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.

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

Chockler, H., Halpern, J.Y., 2004. Responsibility and blame: a structural-model approach. J. Artificial Intelligence Res. 22, 93–115.

Clark, C.W., 1990. Mathematical Bioeconomics: The Optimal Management of Renewable Resources, 2nd edition, Wiley.

Epstein, R.A., 1973. A theory of strict liability. J. Legal Stud. 2 (1), 151–204.

Ferraro, P.J., Sanchirico, J.N., Smith, M.D., 2019. Causal inference in coupled human and natural systems. Proc. Natl. Acad. Sci. 116 (12), 5311–5318.

Friedman, M., 1961. The lag in effect of monetary policy. J. Polit. Econ. 69 (5), 447–466.

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

Good, I.J., 1961. A causal calculus (I). British J. Philos. Sci. 11 (44), 305–318.

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

Granger, C.W., Newbold, P., 1977. Forecasting Economic Time Series. Academic Press.

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

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. Bull. Am. Meteorol. Soc. 97 (1), 99–110.

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

Holland, P.W., 1986. Statistics and causal inference. J. Amer. Statist. Assoc. 81 (396), 945–960.

Holling, C.S., 1973. Resilience and stability of ecological systems. Annu. Rev. Ecol. Syst. 4, 1–23.

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

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

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 Anal.: Int. J. 28 (4), 815–823.

Jordà, Ò., Singh, S.R., Taylor, A.M., 2022. Longer-run economic consequences of pandemics. Rev. Econ. Stat. 104 (1), 166–175.

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

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

Lenzen, M., Murray, J., Sack, F., Wiedmann, T., 2007. Shared producer and consumer responsibility – Theory and practice. Ecol. Econom. 61 (1), 27–42.

Lewis, D., 1973. Causation. J. Philos. 70 (17), 556–567.

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

Mackie, J.L., 1965. Causes and conditions. Am. Philos. Q. 2 (4), 245–264.

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, 1. John W. Parker, London.

Mitroff, I.I., Silvers, A., 2013. Probabilistic causality. Technol. Forecast. Soc. Change 80 (8), 1629–1634.

Mittelstaedt, C., Baumgärtner, S., 2023. Attribution of Collective Causal Responsibility To Individual Actors in a Stochastic System. SSRN Discussion paper 3967091, https://ssrn.com/abstract=3967091.

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

Otto, F.E.L., 2017. Attribution of weather and climate events. Annu. Rev. Environ. Resour. 42, 627–642.

Pearl, J., 2009. Causality. Cambridge 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. Clim. 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. Rev. Econom. Stud. 51 (2), 289–303.

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

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

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

Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66 (5), 688.

Rubin, D.B., 1980. Randomization analysis of experimental data: The fisher randomization test comment. J. Amer. Statist. Assoc. 75 (371), 591–593.

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

Solow, R.M., 1956. A contribution to the theory of economic growth. Q. J. Econ. 70 (1), 65–94.

Stecher, M., Baumgärtner, S., 2022. A stylized model of stochastic ecosystems with alternative stable states. Nat. Resour. Model. 35 (4), e12345. http://dx.doi.org/10.1111/nrm.12345.

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.

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.

Vallentyne, P., 2008. Brute luck and responsibility. Politics Philos. Econ. 7 (1), 57–80.

Van Inwagen, P., 1978. Ability and responsibility. Philos. Rev. 87 (2), 201–224.

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

Wright, S., 1921. Correlation and causation. J. Agric. Res. 20, 557–585.