Research Article

Dominik Janzing* and Sergio Hernan Garrido Mejia

A phenomenological account for causality in terms of elementary actions

https://doi.org/10.1515/jci-2022-0076
received November 16, 2022; accepted October 25, 2023

Abstract: Discussions on causal relations in real life often consider variables for which the *definition* of causality is unclear since the notion of *interventions* on the respective variables is obscure. Asking "what qualifies an action for being an intervention on the variable X" raises the question whether the action impacted all other variables only *through* X or *directly*, which implicitly refers to a causal model. To avoid this known circularity, we instead suggest a notion of "phenomenological causality" whose basic concept is a set of *elementary actions*. Then the causal structure is defined such that elementary actions change only the causal mechanism at *one* node (e.g. one of the causal conditionals in the Markov factorization). This way, the principle of independent mechanisms becomes the defining property of causal structure in domains where causality is a more abstract phenomenon rather than being an objective fact relying on hard-wired causal links between tangible objects. In other words, causal relations between variables get defined by the *interface between the system and an external agent* (who is able to perform the elementary actions), rather than being an *internal* property of links between the variables. We describe this phenomenological approach to causality for toy and hypothetical real-world examples and argue that it is consistent with the causal Markov condition when the system under consideration interacts with other variables that control the elementary actions.

Keywords: causal abstractions, independence of mechanisms, complexity of interventions

MSC 2020: 62H22, 62A09

1 Introduction

While machine learning (ML) plays an increasing role in the technological development of our society (for instance, by providing forecasting systems, search engines, recommendation systems, and logistic planning tools) the ability of ML systems to learn *causal* structures as opposed to mere statistical associations is still at an early stage[1]. A particularly hard task in causal ML is the so-called *causal discovery*, the task of learning the causal graph (including causal directions) from passive observations [2–4]. Given how many problems modern deep learning (DL) has solved via providing computers with massive data [5], one may wonder whether appropriate DL architectures could *learn how to learn* causal structure from data after feeding them with an abundance of datasets with known ground truth. This approach, however, fails alone due to the scarcity of such datasets.[2] This, in turn, raises the question: *Why is there so little benchmarking data with commonly agreed causal structure?* The common answer is that in many real-world systems interventions are impossible,

1 Cf. for instance, [1], page 30: "Some readers may be surprised that I placed present-day learning machines squarely on rung one of the Ladder of Causation..."

*** Corresponding author: Dominik Janzing,** Amazon Research, Tubingen, Germany, e-mail: janzind@amazon.com
Sergio Hernan Garrido Mejia: Amazon Research and Max Planck Institute for Intelligent Systems, Tübingen, Germany, e-mail: shgm@tuebingen.mpg.de

costly, or unethical (e.g. moving the moon to show that its position causes the solar eclipse is costly at least). Although this is a valid explanation, it blurs the question whether the required interventions are *well defined* in the first place. While *defining* interventions on the moon seems unproblematic, it will be significantly harder to agree on a definition for interventions on the gross national product (GNP), for instance, as a basis for discussing the impact of GNP on employment. Which of all hypothetical political instruments (whether feasible or not) influencing GNP should be considered *interventions on GNP* in the sense that GNP is the only variable that is affected *directly*? Likewise, Hernán and Taubman [9] discusses ill-definedness of interventions on body mass index (BMI), which renders discussions of consequences of BMI problematic. Before going into this discussion, we first recall the description of interventions in the framework of graphical models.

1.1 Notation and terminology

Random variables will be denoted by capital letters like X, Y and their values by lower case letters x, y. Further, calligraphy letters like $\mathcal{X}, \mathcal{Y}$ will denote the range of variables X, Y. Causal Bayesian networks (CBNs) or functional causal models (FCMs) [2,3] will be our key framework for further discussions. Both concepts describe causal relations between random variables $X_1, \ldots, X_n$ via directed acyclic graphs (DAGs). According to the causal Markov condition, the respective causal DAG G is compatible with any joint density that factorizes[3] according to

$$p(x_1, \ldots, x_n) = \prod_{j=1}^{n} p(x_j|pa_j), \tag{1}$$

where pa_j denotes the values of the parents PA_j of X_j in G. Then, the CBN is given by a DAG G together with the Markov kernels[4] $p(x_j|pa_j)$.

FCMs, in contrast, provide a deterministic model of the joint distribution, in which every node is a function of its parents and an unobserved noise variable N_j:

$$X_j = f_j(PA_j, N_j), \tag{2}$$

where all $N_1, \ldots, N_n$ are statistically independent. Both frameworks, CBNs and FCMs, admit the derivation of interventional probabilities, e.g. the change of the joint distribution after setting variables to fixed values. While CBNs only provide statements on how probabilities change by an intervention, FCMs, in addition, tell us how the intervention affected each individual statistical unit and counterfactual causal statements [3].

1.2 Interventions

To briefly review different types of interventions, note that the point intervention $do(X_j = x_j)$ [3], also called "hard intervention," adjusts the variable to x_j, while generalized (also called "soft") interventions on X_j replace $p(x_j|pa_j)$ with a different conditional $\tilde{p}(x_j|pa_j)$ or the FCM (2) with a modification $X_j = \tilde{f}_j(PA_j, \tilde{N}_j)$, see the study by Peters et al. [4], page 89, and references therein. Structure-preserving interventions [11] preserve all the dependences on the parents by either operating on the noise N_j or adjusting X_j to the parent-dependent value $f_j(pa_j, N_j')$, where N_j' is an independent copy of N_j that is generated by the experimenter. The simple

2 Even benchmarking the elementary problem of cause–effect inference from bivariate data is often done via the Tübingen dataset [6] containing currently 106 pairs only [7]. More studies are therefore heavily based on simulated data [8].

3 Here, we have implicitly assumed that the joint distribution has a density with respect to a product measure [10].

4 Here, we have slightly abused notation and assumed that $p(x_j|pa_j)$ are defined also for values pa_j with probability density zero, which enables inferring interventional distributions beyond the support of the joint distribution.

observation that any of these interventions on X_j affect only X_j and a subset of its descendants defines a consistency condition between a hypothetical G and the hypothesis that an action is an intervention on X_j.

While the aforementioned framework has provided a powerful language for a wide range of causal problems, it implicitly requires the following two questions to be clarified:

Question 1. (coordinatization) How do we define variables $X_1, \ldots, X_n$ for a given system, that are not only meaningful in their own right but also allow for well-defined causal relations between them?

The second question reads:

Question 2. (defining interventions) Let A be an intervention on a system S whose state is described by the variables $\{X_1, \ldots, X_n\}$ ("coordinates"). What qualifies A to be an intervention on variable X_j only?

Note that the word "only" in Question 2 is meant in the sense that the action intervenes on none of the other variables under consideration *directly*, it only affects them in their role of descendants of X_j, if they are.

Question 1 captures what is often called "causal representation learning" [12,13], while this article will mainly focus on Question 2 only although we believe that future research may not consider them as separate questions for several reasons. First, because some definitions of variables may seem more meaningful than others because they admit more natural definitions of interventions. Second, because different parameterizations of the space results in different *consistency conditions* between variables which may or may not be interpreted as causal interactions between them.

Following the study by Woodward [14], page 98, one may define interventions on X_j as actions that affect only X_j and its descendants, but this way one runs into the circularity of referring to the causal structure for defining interventions although one would like to define *causality via interventions*, see the study by Baumgartner [15] for a discussion. The idea of this article is to define causal directions between a set of variables by first defining a set of *elementary transformations* acting on the system, which are later thought of being interventions on one of the variables only. While they can be concatenated to more complex transformations, defining which transformations are *elementary*, defines the causal direction. While the notion of interventions often appears to be a *primary* concept in the graphical model based framework of causal inference, this article will tentatively describe a notion of intervention as a *secondary* concept derived from an implicit or explicit notion of *complexity of actions*.

1.3 Structure of the article

Section 2 argues that there are domains (mostly technical devices) in which an action can be identified as an intervention on a certain variable via analyzing the "hardware" of the respective system. In contrast, Section 3 describes a few scenarios with ill-defined causal relations to highlight the limitations of the "hardware analysis" approach, without claiming that our proposal will offer a solution to all of them. Section 4 argues that the idea that some operations are more elementary than others is already necessary to make sense of one version to read the principle of independent mechanisms. Based on this insight, Section 5, which is the main part of the article, describes how to define causal directions via declaring a set of transformations as elementary and illustrates this idea for examples of "phenomenological" causality where the causal directions are debatable and paradoxical. Section 6 shows that phenomenological causality appears more natural when the system under consideration is not considered in isolation, but in the context of further variables in the world. Then, the joint system can be described by a DAG that is consistent with phenomenological causality (Subsection 6.1). Further, the notion of "elementary" also satisfies a certain consistency condition with respect to such an extension (Subsection 6.2).

Sections 2 and 3 are not strictly required for understanding the main ideas of this article and can be skipped by readers familiar with the problem of ill-defined causal relations. Section 2 starts with a hypothetical technical toy system in which causal relations and interventions are tangible due to their manifestation as

wires that transmit physical signals. Further, it describes a second toy system, in which causal relations get less obvious although the system is actually tangible, which can be seen as a metaphor and motivation for Section 3, which is a collection of scenarios that can render causal relations ill-defined.

2 Defining interventions via "hardware analysis"

We want to motivate Questions 1 and 2 from the previous section by two thoughts experiments, starting with Question 2. Consider the apparatus shown in Figure 1. It shows n measuring devices that display real numbers $X_1, X_2, \ldots, X_n$. Further, it contains n knobs, whose positions are denoted by $K_1, \ldots, K_n$. Assume we know that the box contains an electrical device and X_j are n different voltages whose mutual influence is described by some unknown DAG. In the best case, our knowledge of how the device is internally wired tells us that turning knob j amounts to intervening on X_j. In the worst case (the black box scenario), our judgement of whether K_j intervenes on X_j is only based on observing the impact on all X_i, where we run in the aforementioned circularity of checking whether K_j only affects X_j and a subset of its descendants. However, depending on "hardware analysis" for defining interventions in a non-circular way is worrisome. First, the power of the causal inference framework relies on the fact that it describes causal relations on a more abstract level without referring to the underlying "hardware." Second, it is questionable why analyzing the causal relation between the action at hand and a variable X_j should be easier than analyzing the causal relations between different X_i (e.g. following wires between the knobs and the voltages X_j as well as the wires between different X_i both requires opening the box). After all, both causal questions refer to the same domain. A large number of relevant causal relations refer to domains with an inherent fuzziness, e.g. macro economic questions, and it is likely that the same fuzziness applies to the causal relation between an action and the variables it is supposed to intervene or not to intervene on. Variables for technical devices like *voltage at a specific component* refer to measurements that are local in space–time and thus require propagating signals to interact, which admits interpreting edges of a causal DAG as those signals. While we do not claim that analyzing the spatial structure of a technical device renders Question 1 and 2 entirely trivial, we do think that spatial structure provides guidance in answering them: choose variables that represent distant parts of a machine (guidance for Question 1), touching one part is clearly not an intervention on variables that correspond to distant parts (Question 2). In contrast, variables in many other domains, e.g. GNP in economics are highly non-local, which renders causal edges an abstract concept.

Part of the fuzziness of causality in "high-level variables" in real-life applications can be captured by the following metaphoric toy example: to motivate Question 1, consider the apparatus show in Figure 2. Instead of showing n measuring instruments on its front side, it contains a window through which we see a mechanical device with a few arms and hinges, whose positions and angles turn out be controlled by the positions of the knobs. The angles and positions together satisfy geometric constraints by construction, and they cannot be changed independently.[5] Accordingly, parameterizing the remaining 4 degrees of freedoms via variables

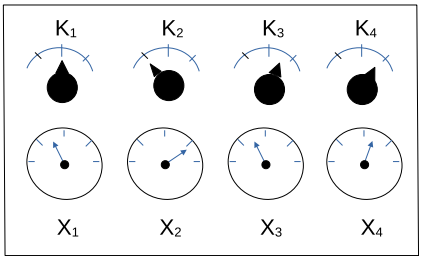

Figure 1: Apparatus whose front side contains n measurement devices and n knobs. The measuring devices measure unknown quantities $X_1, \ldots, X_n$. How do we "find out" (or "define"?) whether knob j intervenes on X_j?.

5 For causal semantics in physical dynamical systems, see the study by Bongers et al. [16].

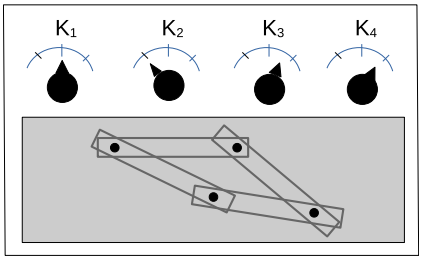

Figure 2: Apparatus whose front side displays n knobs and a mechanical device whose degrees of freedoms we want to parameterize by variables $X_1, \ldots, X_n$. If turning the knobs is the only way to act on the device, different observers may not agree on whether knob i is an intervention on X_j. After all, they can disagree on whether the observed impact on X_j is direct or mediated by some X_i.

$X_1, \ldots, X_4$ is ambiguous, e.g. horizontal and vertical position of one of the 4 hinges and 2 angle (describing the system by more than four variables would be over-parameterization, which results in constraints, as described in Subsection 3.1). Now, the question whether or not turning knob j can be seen as an intervention of one particular mechanical degree of freedom X_j depends on the parameterization. In this case, even hardware analysis does not necessarily clarify which degrees of freedom the knob intervenes on. The example shows also that there is an obvious "cheap" solution to the ill-definedness of interventions: We claim that the position of the knobs are the only variables we are able to intervene on and only explore effects of those, while avoiding questions on causal relations between the *internal* variables. Accordingly, one would deny talking about interactions *between genes* in gene expression experiments and consider changing experimental conditions (the "knobs") as the actual interventions. While this perspective sounds clean and circumvents hard conceptual problems of causality, it dismisses the idea of a causal understanding of the processes *inside the box*.

3 Ill-defined causal relations

We now try an incomplete taxonomy of reasons that render causality ill-defined, many of which remain even after understanding the underlying processes. The main purpose is to show the variety of reasons for ill-definedness as motivation for our ideas, without claiming to provide a coherent picture as solution. By using the apparatus in Figure 2 as a metaphor, we argue that sometimes we still do not understand causal relations even *after opening the box*. This is because causal relations between variables are not always like "wires that link devices." We emphasize that some ill-definedness of causal relations in real-life result from ill-definedness of the variable they refer to, thus leading us to Question 1. For instance, the question to what extent the air temperature outside today influences the value tomorrow: if "temperature today" is supposed to only refer to the temperature of a cubic decimetre around the temperature sensor, the impact is negligible. If, however, it refers to the mean temperature of the *whole region*, the impact is significant. While these temperatures largely coincide for passive observations, interventions destroy this equality and thus intervening on "the temperature" is an ill-defined concept. Let us now discuss a few reasons for ill-definedness of causal relations that appears even when the variables are well-defined. The purpose of providing the below incomplete list of reasons is to argue that discussions of causality often fail to provide *enough context* to get a well-defined causal question. We will later see in what sense this context can be given by specifying those elementary actions that we want to consider *interventions on one variable*.

This ill-definedness of causal relations and the unclarity about what an intervention means for many variables in a causal graph has led several authors to the conclusion that causal claims are supposed to be relative to concrete actions among which an agent can choose, e.g. [9,17]. Our toy examples later suggest, however, that we can give a better causal interpretation to the downstream impact of these actions, if we

endow the set of actions with a notion of complexity and distinguish between "elementary" versus "concatenated" actions. Without ending with a clear conclusion, this section is supposed to encourage the reader to think about the extent to which causality between the variables under consideration is induced by the interaction with the remaining part of the world rather than being an internal property of the system itself.[6] This also aligns with ref. [18], page 66, stating that the definition of causality requires *open* systems.

3.1 Coupling induced by "coordinatization" of the world

For some person, let us define the variables W, E, S, O describing the hours he/she spends for work, exercises, sleep, and others at some day. These variables satisfy the equation

$$W + E + S + O = 24, \tag{3}$$

a constraint which results in statistical dependences when observing the variables over several days. According to Reichenbach's principle of common cause [19], each dependence is due to some *causal* relation, either the variables influence each other or they are influenced by common causes. Since one may be tempted to explain the statistical dependence by a common cause, let us introduce a hidden variables *the person's decision* influencing all four variables, as shown in Figure 3. However, this DAG suggests that independent interventions were possible. Obviously, the intervention of setting all four variables to 7 h is prohibited by (3). Likewise, it is not possible to intervene on W only without affecting the other variables. While these remarks seem obvious, they raise serious problems for defining downstream impact of the above variables, since it is impossible to isolate, for instance, the health impact of increasing W from the health impact of the implied reduction of E, S, or O. Constraints that prohibit variable combinations a priori (by definition), rather than being a result of a mechanism (which could be replaced by others), are not part of the usual causal framework. The fact that independent changes of the variables are impossible also entails the problem of interventions being an ill-defined concept: What qualifies a change of life-style which increased E and decreased W as *an intervention on E*? Is it a question of the intention? If the intention to do more exercises entailed the reduction of work we tend to talk about an intervention on E, but given that people do not always understand even their own motivation behind their actions, it seems problematic that the definition of an intervention then depends on these speculations. Examples like these have motivated the development of more general causal frameworks [20] that deal with such constraints. We will elaborate on the link to our approach in Subsection 5.2.

3.2 Dynamical systems

Constraints on variables like the ones mentioned earlier result naturally if we do not think of the state of the world as a priori given *in terms of variables*. Let us, instead, consider a model where the admissible states of the world are given as points in a topological space (in analogy to the *phase space* of a classical dynamical system [21]), in which variables arise from introducing coordinates, as visualized in Figure 4.

Here, the ellipse defines the admissible states of the world, and we have introduced a coordinate system whose coordinates define the two random variables X^1 and X^2. The states of our toy world are formally given by the set of all those pairs (x^1, x^2) that belong to the ellipse E. We assume that the dynamics of our toy world is given by a topological dynamics of the ellipse, that is, a continuous map that maps each state $p_t \in E \subset \mathbb{R}^2$ to its time evolved state $p_{t+\Delta t}$. The fact that the state cannot leave the ellipse entails a relation between the variables

6 As a reviewer put it, the examples show that the naïve picture on which causes are just read off of the spatiotemporal mechanistic organization is inadequate and why there is therefore room for genuine ambiguity about the causal ordering that needs to be resolved by extrinsic actions.

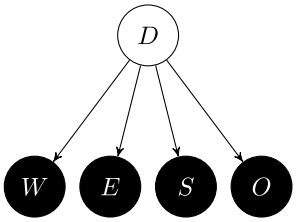

Figure 3: The hours spent with certain "activities" like work, exercise, sleep, and others are determined by a person's decision.

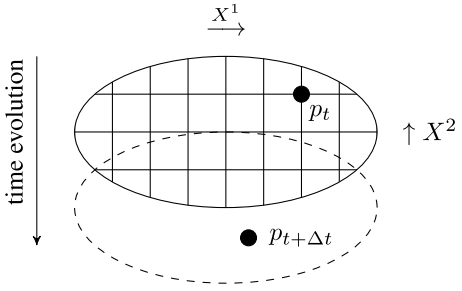

Figure 4: A model of the world where the current state is a point in a topological space, that moves from point p_t to $p_{t+\Delta t}$.

X^1 and X^2 of which it is unclear whether we should consider it a *causal* relation. We will therefore discuss how to define the notion of "interventions" on X^1 and X^2.

It is natural to interpret the intervention set X^1 to $\tilde{x}^1$ as an operation that maps a state (x^1, x^2) to $(\tilde{x}^1, x^2)$. This is, however, only possible if $(\tilde{x}^1, x^2)$ is still a point in the ellipse. Otherwise, one is forced to change x^2 to some other value $\tilde{x}^2$ to still end up in an admissible state of the world. Could we then say that X^2 changed *because* X^1 changed? This interpretation may be valid if someone's intention was to change X^1, who is then forced to also change X^2 by the existing constraints. However, without this knowledge about what has been the intention of the actor, we just observe that both variables have been changed. Interpreting the action as an intervention on X^1 becomes questionable. Our simple model of the world already suggests two different notions of causality:

(1) **Causality between time-like measurements**. Here, we consider the variables X_t^1, X_t^2 as causes of $X_{t+\Delta t}^1, X_{t+\Delta t}^2$.
(2) **Causality without clear time separation**. Here, causal relations between the coordinates appear as a phenomenon emerging from consistency conditions that define the admissible states of the world.

Note that a more sophisticated version of constraints for dynamical systems can arise from equilibrium conditions [22] since the set of active constraints are active only for some interventions, while they get inactive for others [23], which motivates the so-called causal constraints models (CCMs). The fact that X_t^1 and X_t^2 refer to the same points in time suggests to attribute their observed statistical dependences (which result from most distributions on the ellipse) to their common history and draw the causal DAG

$$X^1 \leftarrow H \rightarrow X^2,$$

where H encodes the relevant feature of the state (X_{t-1}^1, X_{t-1}^2). However, this DAG suggests the existence of independent interventions on X_t^1 and X_t^2, although the constraints given by the ellipse need to be respected by any action (similar to our remarks on Figure 3). This fact would rather be covered by a causal chain graph containing an undirected link $X_t^1 - X_t^2$, as discussed in the study by Lauritzen and Richardson [24] for mutual interaction in equilibrium states. If we think of a force in the direction of the x^1-axis and observe that the point moves also in x^2-direction once it reaches the boundary, we would certainly consider X^1 the cause and X^2 the effect. However, once the boundary is reached, it is no longer visible that it's a force in x^1-direction that drives the curved motion. Accordingly, the causal direction becomes opaque.

The second case is more interesting for this article since it refers to a notion of causality that is less understood. At the same time, it challenges the interpretation of causal directions as a concept that is necessarily coupled to time order. Instead, this kind of *phenomenological* causality can also emerge from relations that are not given by the standard physical view on causality dealing with *signals* that propagate through *space–time* from a sender to a receiver[7] (such a physically local concept of causality admits defining interventions on a variable X_j as actions for which a signal from the actor reaches the location of X_j). There is no analogous approach for phenomenological causality, since it may be too abstract to be localized in space–time. For these reasons, Question 2 should be particularly raised for causality in domains outside physics when referring to sufficiently abstract variables. However, variables X^1 and X^2 referring to different coordinate of the same system can also have an abstract causal relation.[8]

3.3 Undetectable confounding: distinction between a cause and its witness

To distinguish between the scenarios $X \to Y$ and $X \leftarrow \tilde{X} \to Y$, where $\tilde{X}$ is latent, is one of the most relevant and challenging problems. Methods have been proposed that address this task from passively observing $P(X, Y)$ subject to strong assumptions, e.g. [26,27] or in scenarios where X, Y are embedded in a larger network of variables, e.g. the fast causal inference algorithm [2] or instrumental variable based techniques [28] and related approaches [29]. Certainly, the task gets arbitrarily hard when the effect of $\tilde{X}$ on X gets less and less noisy, see Figure 5, left. In the limiting case where X is an exact copy of $\tilde{X}$, no algorithm can ever tell the difference between the two scenarios in Figure 5 from passive observations. Note that this scenario is quite common when $\tilde{X}$ is some physical quantity and X, the value of $\tilde{X}$ shown by a precise measurement device. In this case, we would usually not even verbally distinguish between X and $\tilde{X}$ although it certainly matters whether an intervention acts on X or $\tilde{X}$. The distinction gets only relevant when we act on the system, and even then, only if actions are available that decouple X from $\tilde{X}$. Accordingly, it is the set of available actions that defines causality, which is the crucial idea presented in Section 5.

A similar ill-definedness occurs in scene understanding in computer vision: Imagine a picture of a dog snapping for a sausage. Most people would agree that the sausage is the *cause* for the presence of the dog (unless the scene is taken from a location where the dog commonly stays anyway, e.g. its doghouse). However, the presence of the sausage *on the image* is certainly not the cause for the presence of the dog on the image – changing the respective pixels in the image will not affect the pixels of the dog. Similar to the measurement device, we can only meaningfully talk about causal relations if we do not distinguish between the cause "presence of the sausage" and its "witness," e.g. its occurrence on the image. Ignoring actions that decouple the presence of an object in the real scene from its appearance on the image (retouching), we may also consider the presence of the sausage on the image the cause of the presence of the dog on the image and call this causal relation "phenomenological."

3.4 Coarse-grained variables

If coarse-grained ("macrosocopic") variables are defined, for instance, by averaging over "microscopic" variables, interventions on the former are usually no longer well-defined since different changes on the micro-level amount to the same change of the macro-variable [30]. Thus, causal relations between macro-variable may be ill-defined. Accordingly, Rubenstein et al. [31] state a consistency condition for coarse-grainings of causal structures according to which interventions on micro-variables that result on the same intervention on the macro-variables also entail the same downstream impact on other macro-variables. Beckers and Halpern [32]

7 Prohibits instantaneous influence between remote objects since no signal can propagate faster than light [25].

8 Note that the study by Bongers et al. [16] also discusses interventions that change some coordinates in dynamical systems, and asks whether position and momentum of a physical particle allow for separate interventions.

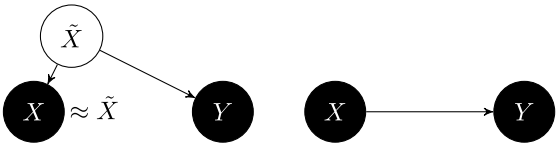

Figure 5: If X is an arbitrarily perfect copy of the confounder $\tilde{X}$ (e.g. if X is a reasonably good measurement of the true quantity $\tilde{X}$), distinction between the left and the right scenario gets arbitrarily hard.

argue that different strengths of consistency conditions are needed for different levels of abstraction. We believe that discussions on causal relations in real life often refer to macro-variables for which the impact of interventions does highly depend on how the intervention is implemented, and it is thus hard to identify *any* valid consistency condition, however weak it may be. Let us consider the following toy model. Given the variables X_1, X_2, Y_1, Y_2, with the causal DAG shown in (4), where X_1 influences Y_1 and Y_2 influences X_2:

$$\bar{X} \begin{Bmatrix} X_1 \longrightarrow Y_1 \\ X_2 \longleftarrow Y_2 \end{Bmatrix} \bar{Y}. \tag{4}$$

We assume FCMs

$$\begin{aligned} Y_1 &= X_1 \\ X_2 &= Y_2, \end{aligned}$$

and define the macro-variables $\bar{X} := (X_1 + X_2)/2$ and $\bar{Y} := (Y_1 + Y_2)/2$, whose causal relations we want to discuss. Obviously, neither an intervention on $\bar{X}$ nor on $\bar{Y}$ is well defined. Increasing $\bar{X}$ by Δ can be done by adding any vector of the form $(\Delta + c, -c)$ with $c \in \mathbb{R}$, and likewise for intervention on $\bar{Y}$. Someone who changes $\bar{X}$ and $\bar{Y}$ by changing X_1 and Y_1, respectively, will claim that $\bar{X}$ influences $\bar{Y}$, while someone changing the macro-variables by acting on X_2 and Y_2 only considers $\bar{Y}$ the cause of $\bar{X}$. This suggests that not only the quantitative effect but also the causal direction is not a property of the system alone (even after specifying the coarse-graining), but a result of which actions are available.

3.5 Diversity of non-equivalent interventions

In the framework of FCMs and graphical causal models, the impact of interventions (e.g. point interventions $\mathrm{do}(X_j = x_j)$) does not depend on *how* this intervention has been performed. Here, the word "how" is meant in the sense of *which mechanism has been used to change* X_j. The description of the mechanism implementing the intervention is not part of the description, which is because the framework implies that different ways of setting X_j to x_j have the same impact. In real-world problems, however, we talk about impact of one variable on another one without specifying the model we refer to, which renders impact of interventions ill defined. Let us elaborate on this in the context of stochastic processes. Let $(X_t)_t$ be a time series of the electricity consumption of a household, where each value is the integral over 1 h. A simple action to reduce X_t at some t could be to convince the resident not to start his/her dish washer at this point in time. Whether this action causes a change of X_s at some later time $s > t$ depends on whether the resident decides to clean the dishes by hand or to just delay the start of the machine. Likewise, the impact of changing the traffic of some road depends on *how* it is performed. In a scenario where the road is made less attractive by a strong speed limit, drivers may take a different route and thus increase the traffic of other roads. Reducing the traffic by offering additional public transport would not have the same effect. Again, one option to disambiguate is to specify the action that defines the causal impact and challenge the belief in absolute causal truth that holds without this additional specification.[9]

9 We will later discuss an approach [20], in which interventions act on *equations* rather than *variables* which accounts for the ill-definedness of the latter.

3.6 Abstract causal mechanisms

While the previous subsection described difficulties with the concept of causality that already arise in clearly defined physical systems, we now discuss an example where the causal mechanisms lie in a more abstract domain. Examples of causal relations for which we believe the simple mechanistic view of causality to be problematic are widespread in the literature. The study by Schölkopf et al. [13], for instance, describes an scenario of online shopping where a laptop is recommended to a customer who orders a laptop rucksack. The authors argue that this would be odd because the customer probably has a laptop already. Further, they add the causal interpretation that buying the laptop is the *cause* of buying the laptop rucksack. We do agree to the causal order but do not believe the *time order* of the purchases to be the right argument. For someone who buys laptop and laptop rucksack in different shops, it can be reasonable to buy the rucksack *first* to safely carry the laptop home (given that he/she already decided on the size of the laptop). We believe, instead, that the recommendation is odd because the *decision* to purchase the laptop is the cause of the *decision* to buy the rucksack. Whether this necessarily implies that the decision has been made earlier is a difficult question of brain research. If they were made in two well-localized, slightly different positions in the brain, one could, again, argue that causal influence can only be propagated via a physical signal (of finite speed). We do not want to further elaborate on this question. The remarks were only intended to show that causal problems in everyday business processes refer to rather abstract notions of causality – for instance, to causality between *mental* states. Particularly in these domains, causality seems highly context dependent.

Economic variables often show several of the aforementioned aspects of ill-definedness coming from aggregation or psychological factors or both. For example, the price of a particular good is not only understood as the price at which one firm sells that good (unless we are in a monopoly market) but instead as an aggregation of prices. Likewise, consumer confidence indices are an aggregation of the beliefs of individual agents. Economic indices tend to be more abstract than their name might suggest.

4 Complexity aspect of independence of mechanisms (IM)

The idea that the conditionals $p(x_j|pa_j)$ in (1) correspond to "independent" mechanisms of the world, has a long tradition in the causality community, see the study by Peters et al. [4], Section 2.2, for different aspects of "independence" and their history. Janzing et al. [33,34] conclude that the different conditionals contain no algorithmic information about each other, and Schölkopf et al. [35] conclude that they change independently across environments and describe implications for transfer learning scenarios. The "sparse mechanism shift hypothesis" [13] assumes that changing the setup of an experiment often results in changes of $p(x_j|pa_j)$ for a small number of nodes X_j.

Here, we want to discuss this independent change from a slightly different perspective, namely, from the one of *elementary* versus *complex* actions. To this end, we restrict the attention to a bivariate causal relation $X \to Y$. According to the interpretation of IM in [35], the causal structure entails that $P(X)$ and $P(Y|X)$ change *independently* across environments. More explicitly, knowing that $P(X)$ changed to $P'(X)$ between training and test data does not provide any information on how $P(Y|X)$ changed. In the absence of any further evidence, it will thus often be reasonable to assume that $P(Y|X)$ remained the same (which is the so-called covariate shift scenario [36]). Likewise, it can also be the case that $P(Y|X)$ changed to $P'(Y|X)$, while $P(X)$ remained the same. However, the scenario that only $P(Y)$ changed and $P(X|Y)$ remained the same or vice versa is rather unlikely. The reason is that this required contrived *tuning* of the changes of the mechanisms of $P(X)$ and $P(Y|X)$. Let us illustrate this idea for a simple example.

Example 1. (ball track) Figure 6 is an abstraction of a real experiment (which is one of the cause–effect pairs in [6]) with a ball track. A child puts the ball on the track at some position X, where it accelerates and reaches a point where its velocity Y is measured by two light barriers. One can easily think of a scenario where $P(X)$ changes without affecting $P(Y|X)$ from datasets to the other one: an older child will tend to choose positions X

that are higher. On the other hand, changing $P(Y|X)$ without affecting $P(X)$ can be done, for instance, by mounting the light barriers at a different position and thus measuring velocity at a later point where the ball already lost some speed. It requires, however, contrived actions to change $P(Y)$ without changing $P(X|Y)$. This would involve both changes of the child's behaviour *and* changes at the speed measuring unit.

Example 1 shows a complexity aspect of IM that we want to build on throughout the article: changing $P(X)$ or $P(Y|X)$ without affecting the other is easy and requires only *one* action. In contrast, changing $P(Y)$ or $P(X|Y)$ without changing the other one of these two objects is difficult for two reasons: first, it requires changes of both, the distribution $P(X)$ of start positions and the conditional $P(Y|X)$ via shifts of the speed measurement. Second, these two actions need to be tuned against each other. After all, those actions on $P(X)$ and $P(Y|X)$ that are easy to implement (e.g. replace the child, shift the mounting of the light barrier) will probably not match together in a way that affects *only* $P(Y)$ but not $P(X|Y)$. In general, if we assume that not all operations on P(Cause) and P(Effect|Cause) are elementary, it may thus take even a *large number* of operations to change only P(Effect) without affecting P(Cause|Effect). Accordingly, for causal DAGs with n nodes, we assume that all elementary operations change at most one conditional $P(X_j|PA_j)$, but we do not assume that any change of a single conditional is elementary. Further, we do not even assume that *any arbitrary* change of $P(X_j|PA_j)$ can be achieved by concatenations of elementary actions.

To relate this view to known perspectives causal counterfactuals, note that Lewis [37] defined the impact of an event E via a hypothetical world that is most similar to the true one except for the fact that E did not happen, as opposed to a world in which E happened, but also several subsequent actions were taken so that the world gets back to the path it would have followed without E. In the spirit of our article, we could think of E as generated by one elementary action and read Lewis' view as the statement that after one elementary action the world is closer in Lewis' sense to the original one than after several interventions that undo the downstream impact of the first one.

5 Defining causal directions via elementary actions

Here, we describe the main idea of the article which uses the notion of "elementary action" as first principle, and then we discuss quite diverse toy examples. Some of them are directly motivated by practically relevant real-life applications, but we also discuss strongly hypothetical scenarios, only constructed with the purpose of challenging our intuition on causality.

5.1 The bivariate case

To avoid the above circularity of defining interventions in a way that relies on the concept of causality and the other way round, we suggest the following approach:

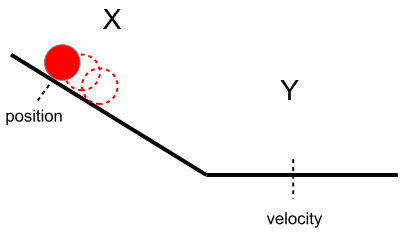

Figure 6: Cause–effect pair with ball-track, taken from dataset [6]: The cause X is he start position of the ball along the inclined plane and the effect Y is the speed at which the ball passes the light barriers at the horizontal track. The example illustrates that $P(X)$ and $P(Y|X)$ correspond to independent mechanisms.

Idea 1. (phenomenological cause–effect pair) Let X and Y be two variables describing properties of some system S and $\mathcal{A}$ be a set of elementary actions on S. We say that X causes Y whenever $\mathcal{A}$ contains only the following two types of actions:

$\mathcal{A}_1$: actions that change X, but preserve the relation between X and Y

$\mathcal{A}_2$: actions that preserve X, but change the relation between X and Y.

Since Idea 1 is quite informal, it leaves some room for different interpretations. We will work with two different ways of spelling it out:

Definition 1. (Statistical phenomenological causality) Let X and Y be two variables describing properties of some system S and $\mathcal{A}$ be a set of elementary actions on S. We say that X causes Y relative to $\mathcal{A}$ whenever $\mathcal{A}$ contains only the following two types of actions:

$\mathcal{A}_1$: actions that change $P(X)$, but preserve $P(Y|X)$ and

$\mathcal{A}_2$: actions that preserve $P(X)$, but change $P(Y|X)$.

Definition 2. (Unit level phenomenological causality) We say that X causes Y relative to $\mathcal{A}$ whenever $\mathcal{A}$ contains only the following two types of actions:

$\mathcal{A}_1$: A set of actions (containing the identity) such that every pair (x', y') obtained from the observed pair (x, y) by an action in $\mathcal{A}_1$ satisfies the same law $y' = m(x')$ for some (non-constant) function m.

$\mathcal{A}_2$: Actions that keep x, but change m. In other words, the state of the system is described by a pair (x, m) where actions in $\mathcal{A}_1$ change x and actions in $\mathcal{A}_2$ change m.

Here and throughout the article, we assume actions to be implemented by external actors and not influenced by the state of the system. Further, we assume acyclic causal structures and exclude confounding.

Note that m in Definition 2 holds for all actions in $\mathcal{A}_1$, but different functions m hold for different statistical units. If we think of X and Y as related by the FCM $Y = f(X, N)$, we should think of m as the map $f(.,n)$ with fixed noise value n. The statement that actions in $\mathcal{A}_1$ do not change the mapping from X and Y thus refers to the counterfactual knowledge encoded by the FCM, which we assume to be given from domain knowledge about the system[10]. Further note that changing the map m can be done by either changing f or n. Although the condition for $\mathcal{A}_1$ is asymmetric with respect to swapping X and Y since m maps from X to Y, this asymmetry does not necessarily imply the mapping m to be *causal*. Assume, for instance, X and Y are related by the FCM $Y = X + N$. Then, an observed pair (x, y) for which $N = 3$ will obey the rule $y = x + 3$ and $y' = x' + 3$ for all pairs generated by actions in $\mathcal{A}_1$. However, all these pairs will also obey the rule $x' = y' - 3$, and thus, there exists also a map $\tilde{m}$ from Y to X. In our following examples, the crucial asymmetry between cause and effect will often not be induced by the existence of m (since a function mapping y to x may also exist), but by the existence of actions $\mathcal{A}_2$, which only act on the effect. In other words, interventions on the cause do not reveal the asymmetry because they change cause and effect, while actions on the effect only change the effect. The idea that actions change equations is just rephrasing the usual interpretation of FCMs as modular components of a system [3]. This view is even more explicit in the study by Spirtes and Scheines [20], which builds upon the study by Simon et al. [38], where interventions are not a priori thought of acting on *variables*. Instead, interventions act on *equations*, which are not even written as assignments (with the effect on the left-hand side), and the causal order then emerges from the direction in which the system of equation is solved. We will comment more on this relation at the end of Section 5.

As an aside, note that for the scenario where X and Y are only connected by a confounder, we would postulate actions that affect only X and those that affect only Y.

The idea of identifying causal structure by observing which conditionals in (1) change independently across datasets can already be found in the literature, e.g. the study by Zhang et al. [39] and references therein. In the same spirit, Definitions 1 and 2 raise the question whether they define the causal direction uniquely. This

10 We think this is justified because the scope of this article is to discuss how to define causality, not how to infer it.

is easier to discuss for Definition 1. Generically, changes of $P(X)$ results in simultaneous changes of both $P(Y)$ and $P(X|Y)$, which thus ensures that the available actions of class $\mathcal{A}_1$ neither fall into the category corresponding to $\mathcal{A}_1$ nor $\mathcal{A}_2$ for the backwards direction from $Y \to X$. The following simple result shows a genericity assumption for which this can be proven:

Proposition 1. (Identifiability via changes) *Let X and Y be finite with finite ranges $\mathcal{X}$ and $\mathcal{Y}$ of equal size, i.,e., $|\mathcal{X}| = |\mathcal{Y}|$, and the square matrix $p(x,y)_{x,y}$ have full rank with $p(x,y)$ strictly positive. Then changing $p(x)$ while keeping $p(y|x)$ changes $p(y)$ and $p(x|y)$.*

Proof. Define $\tilde{p}(x,y) \coloneqq \tilde{p}(x)p(y|x)$. Assume $\tilde{p}(x|y) = p(x|y)$. Hence,

$$\tilde{p}(x)p(y|x)\tilde{p}^{-1}(y) = p(x)p(y|x)p^{-1}(y),$$

which is equivalent to $\tilde{p}(x)p(y) = p(x)\tilde{p}(y)$. Summing over y yields $\tilde{p}(x) = p(x)$ hence $p(x)$ did not change. We conclude that changing $p(x)$ changes $p(x|y)$. That changing $p(x)$ also changes $p(y)$ follows from the full rank assumption. □

Abstract toy example. We now describe an example for unit level phenomenological causality according to Definition 2 where the causal direction is a priori undefined, but may be defined after specifying the set of actions, if one is willing to follow our approach. We have purposefully chosen an example whose causal interpretation seems a bit artificial.

Example 2. (urn model) Assume we are given an urn containing blue and red balls, as well as a reservoir containing also blue and red balls. The game allows four basic operations: (A_1^+) replacing a red ball in the urn with a blue one, (A_1^-) replacing a blue ball with a red one, (A_2^+) adding a red ball to the urn, and (A_2^-) removing a red ball from the urn (and adding it to the reservoir), see Figure 7.

Define the random variables K_b and K_r, describing the number of blue and red balls in the urn, respectively. Obviously, the four different operations correspond to the following changes of K_b and K_r:

$$\begin{aligned}
(A_1^+) &\quad K_b \to K_b + 1; \quad K_r \to K_r - 1\\
(A_1^-) &\quad K_b \to K_b - 1; \quad K_r \to K_r + 1\\
(A_2^+) &\quad K_r \to K_r + 1\\
(A_2^-) &\quad K_r \to K_r - 1.
\end{aligned}$$

Note that action A_2^+ is always possible, but the other three operations are only possible if the quantity to be reduced is greater than zero. According to Definition 2, we have the causal relation $K_b \to K_r$ because the actions $A_1^\pm$ belong to the category $\mathcal{A}_1$ since they preserve the relation $K_b = c - K_r$ for some state-dependent constant c. Further, $A_2^\pm$ belong to the category $\mathcal{A}_2$. We also observe that changing K_b without changing K_r requires the concatenation of two operations at least: For instance, add a red ball, and then "convert" it into a blue one.

The example may be considered as representing a chemical process with two *isomers*, that is, molecules that are structurally different, although they have the same molecular formula. Isomer "red" can be converted into isomer "blue" and vice versa, catalyzed by appropriate experimental conditions [40]. Then, the "red" ones are the resource for a reaction that converts "red" into "blue." Therefore, one may be surprised that not the amount of the *resource* K_r, but the amount of the *product* K_b of the reaction, is the cause.

To come up with another real-life example that resembles that mentioned earlier, let each red ball represent 1 Euro in a transactions account and a red ball be 1 Euro in the savings account. The actions

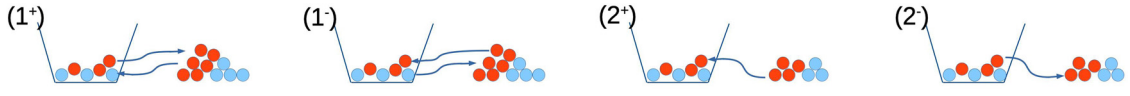

Figure 7: Urn model in Example 2 with four different operations.

correspond with typical actions that a bank account holder has: add/remove money from the transactions account, and move money from transactions to savings account or *vice versa*. We conclude that the phenomenological causal relation between the two is that the amount of money on the savings account causes the amount of money on the transactions account.[11]

We now rephrase Example 2 into an example for the statistical version in Definition 1. To this end, we consider a random experiment for which the system is initially in the state $K_r = k_r$ and $K_b = k_b$ with $k_r, k_b \gg 0$. Then, in each round, we flip a coin for each of the four actions to decide whether they are applied. After $\ell < k_r, k_b$ many rounds, let N_1 denote the number of times A_1^+ minus the number of times A_1^- has been applied. Likewise, N_2 counts the number of A_2^+ minus the number of A_2^- actions. We then obtain

$$K_b = k_b + N_1 \tag{5}$$

$$K_r = k_r - N_1 + N_2. \tag{6}$$

One easily checks that (5) and (6) are equivalent to

$$K_b = k_b + N_1 \tag{7}$$

$$K_r = -K_b + k_r + k_b + N_2. \tag{8}$$

Since the actions are controlled by independent coin flips, we have $N_1 \perp\!\!\!\perp N_2$. Following our interpretation that K_b causes K_r and $A_1^\pm$ and $A_2^\pm$ are interventions on K_b and K_r, respectively, we thus consider (7) and (8) as the corresponding FCM. By controlling actions $A_1^\pm$ and $A_2^\pm$ via coins with different bias, we may change the distributions $P(K_r|K_b)$ and $P(K_b)$ independently, and thus have an example of statistical phenomenological causality in Definition 1.

The following observation may seem paradoxical at first glance: The set $\mathcal{A}_1 = \{A_1^+, A_1^-\}$ is a priori symmetric with respect to swapping the roles of blue and red. The justification for calling it interventions on K_b is derived from properties of $\mathcal{A}_2 = \{A_2^+, A_2^-\}$. In other words, whether an action is considered an intervention on a certain variable depends on the impact of the *other* actions in the set of elementary actions. This context dependence of the definition of interventions may be worrisome, but in scenarios where "hardware-analysis" does not reveal (or define) whether an action is an intervention on a particular variable, we do not see a chance that circumvents this dependence on other actions. Given the abstractness of the underlying notion of causal direction, we are glad to observe that the well-known causal discovery method LiNGAM [41] would also infer $K_b \to K_r$ because (7) and (8) define a linear model with non-Gaussian additive noise. The crucial assumption inducing the statistical asymmetry is that actions $A_1^\pm$ are implemented independently of $A_2^\pm$, resulting in independent noise variables N_1, N_2.

There is also another aspect of this example that shows the abstractness of the causal interpretation of the above scenario. The fact that actions in $\mathcal{A}_1$ preserve the total number of balls has been interpreted as structural equation (8) generating K_r from K_b. Using the function m from Definition 2, this structural equation reads $K_r = m(K_b)$ with $m(K_b) = -K_b + k_b + k_r + N_2$. Since operations in $\mathcal{A}_2$ change the total number of balls, they change m to m' by changing k_b. Hence, actions in $\mathcal{A}_2$ change the "mechanism" relating K_b and K_r. In a mechanistic interpretation of causality, one would expect changes of a mechanism a change of a kind of machine where the input–output behaviour is changed. As abstract as the "mechanism" from K_b to K_r is, as abstract is its change.

Context-dependent causal directions Here, we describe a system for which causal directions swap when the system moves from one regime to another one. Although the following example is hypothetical, we encourage the reader to think of similar example in realistic business processes. While context dependence of causal directions is not a novel insight [20,42,43], we describe the example to show that it is here due to the context dependent impact of elementary actions.

11 We are grateful to an anonymous reviewer for this example.

Example 3. (food consumption of rabbits) Given a hutch with n rabbits where we define two variables:

X: total amount of food consumed by all rabbits at one day
Y: food per rabbit consumed at one day.

By definition, we have $Y = X/n$. We allow the following three types of actions:
$\mathcal{A}_r$: change the number n of rabbits
$\mathcal{A}_f$: change the amount of food provided
$\mathcal{A}_a$: give an appetizer to the rabbits.
We then consider two complementary scenarios, see Figure 8:

Scenario 1: there is enough food for each rabbit. Offering more food influences neither X nor Y. Adding or removing rabbits changes X, but not Y, while the appetizer changes both X and Y and preserves the equality $X = n \cdot Y$. We thus have actions influencing both (while preserving their relation) and those influencing only X, but no action influencing only Y. We thus conclude $Y \to X$.

Scenario 2: shortage of food. Now, the appetizer has no effect. Changing the number of rabbits changes the food per rabbit, but not the total consumption. Changing the amount of food changes consumption per rabbit and total consumption. Hence, we have actions that influence Y, but not X, and actions that influence both, but preserve the relation $Y = X/n$ Further, we have no action influencing X without affecting Y. We thus conclude $X \to Y$.

The aforementioned dependence of the causal direction on the regime suggests that there exists a large grey zone without clearly defined causal direction. We want to discuss this for the following business relevant example, where we will not offer a clear answer.

Example 4. (revenue and sold units) Let R and Q be random variables denoting the revenue and the number of sold units for a company, respectively. Its instantiations are r_j, q_j denoting revenue and number of units for product j. If p_j denotes the price of product j, we thus have

$$r_j = p_j \cdot q_j. \tag{9}$$

If we consider n_j as instantiations of a random variable N, we thus write

$$R = P \cdot Q. \tag{10}$$

One may want to read (10) as structural equation, which suggests the causal structure $Q \to R$. However, this requires P to be independent of Q as a minimal requirement (unless we are talking about a confounded relation). This independence is often violated when more items are sold for *cheap* products. On the other hand, we cannot expect P to be independent of R either, hence neither (10) nor $Q = R/P$ should be considered a structural equation.

To argue that Q causes R one may state that a marketing campaign can increase the revenue by increasing the number of sold units. However, why is a marketing campaign an intervention on Q rather than on R if it increases both by the same factor?

While our intuition may consider $Q \to R$ as the "true" causal direction, the following scenario challenges this. Assume there are two farmers, farmer P producing potatoes and farmer E producing eggs. They have

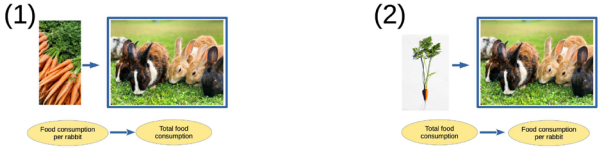

Figure 8: In scenario 1, the food consumption per rabbit is causing the total food consumption because changing the number of rabbits only changes the latter if there is enough food for each rabbit. In scenario 2 with food shortage, changing the number of rabbits does not affect the total food consumption, and it only changes the food consumption per rabbit. Therefore, the total food consumption is the cause. Images by Tom Paolini (carrots), michealcopley03 (single carrot), Aswathy (rabits) with unsplash license.

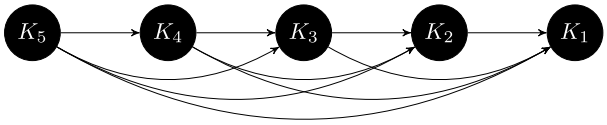

Figure 9: Causal relation between the variables K_j, which count the number of balls in the urn with label j (according to our definition of phenomenological causal structure).

implemented a countertrade with exchanging K_P and K_E with many potatoes and eggs, respectively, according to the negotiated exchange factor F. We then have

$$K_E = K_P \cdot F. \tag{11}$$

For farmer P, K_P is the number of units sold, while K_E is the revenue for her, while farmer E considers K_E the number of units sold and K_P the revenue. If number of units is always the cause of the revenue, then the causal direction depends on the perspective. Preference for one causal direction versus the other could comes from insights about which quantity reacts more to changes of F: assume, for instance, the number of potatoes exchanged is more robust to changes of F (i.e. in economic terms, the demand of potatoes has small price elasticity), then we could consider changes of F as interventions on K_E and thus conclude $K_P \to K_E$.

Example 4 shows that causal directions can also be in the grey zones because *actions* can be in the grey zone of being an intervention on one versus the other quantity. Acting on the factor F will in general change both variables R and Q. However, in the regime where one of it is relatively robust to changes of F, we can consider this one as the cause and consider changing F as an intervention on the effect because it affects the mechanism relating cause and effect. It is likely that many causal relations in real life show equally much room for interpretations.

Let us now revisit the example from the domain of product recommendation algorithms described in the study by Schölkopf et al. [13]:

Example 5. (Laptop and its rucksack) Let X, Y be binaries that describe the decisions of a person to buy or not to buy a laptop and a laptop rucksack, respectively. Let $P(X, Y)$ be the prior joint distribution without any marketing campaign. Let actions in $\mathcal{A}_l$ be marketing campaigns that try to sell more laptops (without explicitly mentioning laptop rucksacks). Then it is likely that these actions influence $P(X)$, but not $P(Y|X)$. Let actions $\mathcal{A}_r$ define marketing campaigns that target at selling laptop rucksacks. Let us assume that this changed $P(Y|X)$, but not $P(X)$. Then we may conclude that X causes Y relative to these sets of elementary actions.

Here, we have neglected that seeing laptop rucksacks may remind some customers that they were planning to buy a laptop already since a while, which could induce additional demand for laptops. Further, a marketing campaign changing $P(X)$ and $P(Y|X)$ could certainly exist, e.g. one that explicitly advertises laptop and rucksack as an economically priced pair. We are thus aware of the fact that any causal statement in this vague domain of customer psychology is a good approximation at best. After all, calling the aforementioned operations "elementary" comes from the idea that they appear quite natural to us, but we admit that this view is debatable. When previously mentioning the scenario of Example 5 in the introduction, we have emphasized that the *time order* of customer's purchases does not determine the *causal order* of the underlying decisions for the purchases. This already suggests that analyzing the causal order of the underlying materialized processes does not reveal the causal structure of their psychologic origin. The following example elaborates on this discrepancy of phenomenological causal direction and causal direction of underlying micro-processes.

Example 6. (vending machine) In contrast to usual purchasing process, a vending machine outputs the article clearly *after* and *because* the money has been inserted. Accordingly, inserting the money is the cause of obtaining the product. We will call this causal relation "microscopic." For some cigarette vending machine, let X be the number of packages sold at a day and Y be the total amount of money inserted at the same day.

Our microscopic causal relations suggests to consider Y the cause of X, but previous remarks on the relation between revenue and number of sold units suggest the opposite. Let us therefore ask for "natural actions" on the system. Assume we stop some of the smokers on their way to the machine and convince them not to buy cigarettes. This clearly impacts both X and Y. Another action would be to slightly change the price of the packages by manipulating the vending machine. If the change is small enough, it will only affect Y but not X. We thus have a natural action influencing both and one influencing only Y, which suggest that X influences Y, in agreement with what we said about revenue and sold units, but in contrast to the *microscopic* causal structure.

5.2 The multivariate case

We first generalize Definition 1 to multiple variables. Although these generalizations are straightforward, we will see that our multivariate extensions of the urn example reveal the abstractness of phenomenological causality even more.

Definition 3. (multivariate causality, statistical) Let $\mathcal{A}$ be elementary actions on a system described by the variables $X_1,\dots, X_n$. Then we say that G is a valid causal graph relative to $\mathcal{A}$ if $P(X_1, \dots,X_n)$ is Markov relative to G and $\mathcal{A}$ consists of non-empty classes $\mathcal{A}_1,\dots, \mathcal{A}_n$ such that actions in $\mathcal{A}_j$ change $P(X_j|PA_j)$ but no conditional $P(X_i|PA_i)$ for $i \neq j$.

Note that Definition 3 may admit a large number of DAGs. However, there are cases where it uniquely determines a causal order, for instance if all actions in $\mathcal{A}_j$ change the distributions of X_i for all $i \geq j$. Given the causal order, the causal DAG can be determined subject to minimality[12]

Likewise, we generalize Definition 2:

Definition 4. (multivariate causality, unit level) Adopting the setting from Definition 3 we say that G is a valid causal graph relative to $\mathcal{A}$ if $\mathcal{A}$ decomposes into classes $\mathcal{A}_j$ such that for every statistical instantiation $(x_1, \dots,x_n)$, there are maps $m_1,\dots, m_n$ with

$$x_i = m_i(pa_i), \tag{12}$$

such that actions in $\mathcal{A}_j$ preserve all equations (12) valid for $i \neq j$. In other words, the state of the system is described by n "mechanisms" $(m_1, \dots,m_n)$, where actions in $\mathcal{A}_j$ change the jth component.

We now generalize Example 2 to n different balls, where the causal structure suggested by our definition gets even less obvious:

Example 7. Given n balls with labels $j = 1,\dots, n$. Given the actions A_j^+ and A_j^-, which replace one ball of type $j-1$ with j for $j = 2,\dots, n$ or vice versa, respectively. Further, $A_1^\pm$ are defined as adding or removing balls of type 1. If k_j^0 denotes the initial number of balls of type j, and N_j denotes the number of actions A_j^+ minus the number of A_j^-, the number K_j of balls is given by

$$K_j = k_j^0 + N_j - N_{j-1} \quad \text{for } j \geq 2 \tag{13}$$

$$K_1 = k_n^0 + N_1. \tag{14}$$

12 That is, the joint distribution is not Markovian to any proper subgraph [44], which results in a unique DAG for strictly positive distributions.

Let us first recalibrate K_j to $\tilde{K}_j := K_j - k_j^0$. We then introduce vectors $\mathbf{k}^0 := (k_1^0, \ldots, k_n^0)$ and vector valued variables $\tilde{\mathbf{K}} := (\tilde{K}_1, \ldots, \tilde{K}_n)^T$, $\mathbf{N} := (N_1, \ldots, N_n)^T$. Using the Töplitz matrix S with diagonal 1 and second diagonal -1 (and zero elsewhere), we can rewrite (13) and (14) as follows:

$$\tilde{\mathbf{K}} = S\mathbf{N}. \tag{15}$$

This, in turn, can be rewritten as follows:

$$\tilde{\mathbf{K}} = A\tilde{\mathbf{K}} + \mathbf{N}, \tag{16}$$

with the lower triangular matrix

$$A := I - S^{-1} = \begin{pmatrix} 0 & & \cdots & & 0 \\ -1 & 0 & & & \\ \vdots & -1 & \ddots & & \vdots \\ & & & & \\ -1 & & \cdots & -1 & 0 \end{pmatrix}. \tag{17}$$

Equivalently, we can then rephrase (15) by the structural equations

$$\tilde{K}_j = \sum_{i>j} - \tilde{K}_i + N_j. \tag{18}$$

The causal structure for the K_j, which is the same as for $\tilde{K}_j$, is shown in Figure 9. Note that there is exactly one structure matrix A that admits writing each K_j as a linear expression of some K_i and N_j such that A is lower triangular for some ordering of nodes. This is because S uniquely determines A. Assuming linear structural equations, we thus obtain Figure 9 as the unique DAG corresponding to the defined set of elementary actions.

Note that the algebraic transformations between (15) and (17) resemble the algebra in independent component analysis (ICA)-based multivariate causal discovery [45] (following the idea of LiNGAM [41] mentioned for the bivariate case mentioned earlier). This analogy is not a coincidence: ICA decomposes the vector $\mathbf{K}$ into independent noise variables $\mathbf{N}$. Accordingly, since (16) is a linear acyclic causal model with *independent non-Gaussian* noise variables N_j, multivariate LiNGAM would also identify the same causal structure and FCMs that we derived as phenomenological causal model. In other words, if we ensure that the choice of the actions is controlled by random generators, independently across different $\mathcal{A}_j$, we obtain a joint distribution $P(K_1, \ldots, K_n)$ for which the causal discovery algorithm LiNGAM infers the DAG in Figure 9.

It is instructive to discuss Example 7 from the perspective of complexity of some actions that are *not* elementary. Increasing K_j without affecting the others requires j operations, e.g. one can first increase K_1 and propagate this increase to K_j. From the causal perspective, these actions are necessary to compensate the impact of K_j on its child.

A further remark on causal faithfulness [2]. The fact that an intervention only propagates to the child, but not to the grandchild shows that the structural equations are non-generic; direct and indirect influence of K_j on K_{j-2} compensate. Accordingly, if we control each action by independent coin flips as in the remarks after Example 2, the induced joint distribution will not be faithful to the causal DAG. The idea of "nature choosing each mechanism $p(x_j|pa_j)$ in (1) independently" seems to have its limitation here. The reason is that the actions $A_j^{\pm}$ are the building blocks of the system, rather than the Markov kernels $p(x_j|pa_j)$, which are constructed from the former. There is also another "paradox" of our causal interpretation that becomes apparent for $n > 2$, while it seems less paradoxical in Example 2: imagine what happened if we were to redefine $A_0^{\pm}$ as adding or removing of balls of type n instead of type 1. We would then reverse all the arrows in Figure 9. In other words, the direction of the arrows in a long chain of variables depends on the set of available actions *at the end points*.[13] This idea is in stark contradiction to the spirit of modularity [46] assuming each $p(x_j|pa_j)$ is an independent mechanism of nature. The reader may see this as an indicator against interpreting the equations

13 An anonymous reviewer mentioned another scenario in which causal directions in a chain also depend on end points, namely, a chain of heat reservoirs, where heating one end results in propagation through the chain. This example, however, requires actions at the end points to be factual, while here the causal direction depends on which actions are *possible*.

(18) as FCMs, but we think that causal directions on the phenomenological level may well depend on this kind of *context*.[14] One option to restore modularity is to consider the equations themselves as mechanisms (for the cost of having no direction a priori), as suggested in the study by Blom et al. [20].

In Example 7, the locality of the impact of each of the actions $A_j^{\pm}$ itself (affecting only two adjacent variables) entailed long-range causal influence between the variables. Now we will describe the opposite where actions affecting a large number of variables is induced by only *local* causal connections (in other words: in the first example, S has only entries in the first off-diagonal, in the case following now this is true for A).

Example 8. (n different balls in bundles) We now modify Example 7 such that the n balls come in the following bundles: there are n different types of packages and type P_j contains the balls $1,\dots,j$ (one per package). Then there are $2n$ different actions $A_1^+, A_1^-, \dots, A_k^+, A_k^-$ of the following form: A_j^+ puts one package P_j from the stack into the urn, while A_j^- wraps balls with label $1,\dots,j$ to one package and puts them back to the stack. We then introduce n random variables, $K_1,\dots,K_n$, where K_j is the number of balls with label j in the urn. Obviously transformation A_j^+ simultaneously increases all the variables $K_1,\dots,K_j$ by 1, while A_j^- decreases all of them by 1, as depicted in Figure 10 for $n = 4$.

By using the same derivation and notation as in Example 7, we define N_j as the difference of actions $A_j^{\pm}$ and obtain

$$\tilde{K}_j = \sum_{i \geq j} N_i, \tag{19}$$

which yields $\tilde{\mathbf{K}} = S\mathbf{N}$ with

$$S \coloneqq \begin{pmatrix} 1 & 0 & \cdots & 0 \\ & 1 & 0 & \\ \vdots & & \ddots & \vdots \\ 1 & & \cdots & 1 \end{pmatrix}.$$

For the structure matrix, we thus obtain the lower triangular matrix

$$A = I - S^{-1} = \begin{pmatrix} 0 & & \cdots & & 0 \\ 1 & 0 & & & \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

which amounts to the structural equations

$$K_n = N_n, \tag{20}$$

$$K_j = K_{j+1} + N_j \quad \forall j \leq n-1. \tag{21}$$

These equations correspond to the causal DAG in Figure 11.

An intervention that changes K_j necessarily changes all K_s with $s < t$ by the same amount, as a downstream impact, according to (21). While the transformations $A_j^{\pm}$ change all K_s with $s < t$ per definition, it is a priori not obvious to see which of these changes should be considered direct and which one indirect. However, the causal interpretation (21) clearly entails such a distinction.

What's the purpose of the causal interpretation? The balls in the urn show an extreme case where the causal interpretation is far away from any "mechanistic view" of causality where the functions m_j from Definition 4 refers to tangible mechanisms (recall, for instance, that $\mathcal{A}_j$ in Example 7 were symmetric with respect to swapping j and $j-1$, yet we have identified them as interventions on K_j, not on K_{j-1}). To argue that

14 Note that this dependence of causal directions on end points also happens when causality is defined by causal ordering [20].

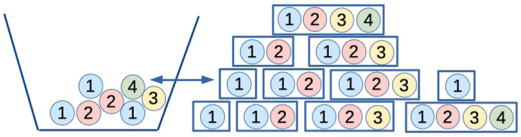

Figure 10: Urn containing packages of n different types, where type P_j contains balls will label $1,\dots,j$. The variable K_j counts the number of balls with symbol j in the urn. The elementary operations of the system are adding one package from the stack to the urn or put it back. Changing K_j thus entails the same change for $K_{j1},\dots,K_1$.

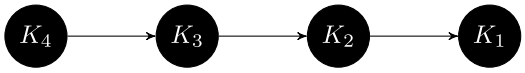

Figure 11: Causal relation between the variables K_j, which count the number of balls in the urn with label j (according to our definition of phenomenological causal structure).

our causal interpretation is not just a weird artefact of our concept, we need to show its benefit. The following section will argue that this extension of causality allows us to consistently talk about the overall causal DAG when systems with "phenomenological causality" are embedded in systems with more tangible causal structure.

Related ideas in the literature. While the idea that causal conditionals $p(x_j|pa_j)$ and structural equations define "modules" in a causal DAG which can be manipulated independently has a long tradition, we have turned it somehow around by defining these manipulations as the primary concept and the causal DAG (in case the set of elementary actions correspond to a DAG, see below) as a derived concept. The closest work to this idea seems to be the study by Blom et al. [20], where DAGs also appear as derived concepts rather than being primary. The idea is to start with a set of equations, where each can contain containing endogenous variables as well as exogenous ones. Subject to certain conditions (phrased in terms of matchings in bipartite graphs), one can uniquely solve the equations to express the endogenous variables in terms of the exogenous ones using Simon's ordering algorithm [38]. Remarkably, general interventions in the study by Blom et al. [20] are thought to act on *equations* rather than *variables*. Note that in the usual view of causality, an action that changes the structural equation $X_j = f_j(PA_j, N_j)$ to some different equation $X_j = \tilde{f}_j(PA_j, N_j)$ is considered an intervention *on* X_j only because the equation is read as a "structural equation" (or "assignment") for X_j, rather than for any of the parents or the noise N_j. In previous studies [20,38], an equation is not *a priori* considered an assignment for a certain variable, but only later after analyzing the direction in which the system of equations is solved. We now argue that our view complements this view. To this end, we consider the following set of equations:

$$X_1 = N_1 \tag{22}$$

$$X_2 = \alpha X_1 + N_2. \tag{23}$$

Assume N_1 and N_2 are independent, as in unconfounded structural equations. When interpreted as purely mathematical equations, they are equivalent to

$$X_2 = \tilde{N}_1 \tag{24}$$

$$X_1 = \tilde{\alpha} X_2 + \tilde{N}_2, \tag{25}$$

when we define appropriate noise variables $\tilde{N}_1, \tilde{N}_2$. For Gaussian N_1, N_2, they can still be independent if $\tilde{\alpha}$ is properly chosen. The crucial questions now read: First, what is the justification for reading equations (22) and (23) as causal mechanisms as opposed to (24) and (25)? Second, what is the justification to prefer N_1, N_2 as exogenous noise instead of $\tilde{N}_1, \tilde{N}_2$? To start with an unsatisfactory answer, one could say that N_1, N_2 (or, more precisely $P(N_1), P(N_2)$) or the equations vary independently across environments, while $\tilde{N}_1, \tilde{N}_2$ would not, but this raises the question which notion of independence such a statement should refer to. If it is meant to be

statistical independence, we would be concerned about building causal semantics on a somewhat arbitrary probability distribution over environments.

A different answer is given by Blom et al. [20] building on the study by Simon [38], where causal order comes from a set of equations which need to be solved in a certain order such that every variable is uniquely determined by previous ones up to a exogenous noise. Although each single equation does not show an asymmetry between cause and effect because it is a priori not written as an assignment with effect on the left-hand side (in contrast to our aforementioned equations), the asymmetry is encoded in the *system* of equations, where actions change only *one* equation (or one noise variable, see footnote 4 in the study by Blom et al. [20]) in the set.[15] However, we could concatenate this type of intervention also in a way that they change only *one equation* of the derived equations (24) and (25), or *one of the noise terms* in the latter. Therefore, we assume that the preference for any among these equivalent set of equations in the study by Blom et al. [20] comes from an implicit notion of *elementary* versus *concatenated* actions.

To make this unified view on the study by Blom et al. [20] and ours explicit for Example 2, we can justify our choice of exogenous variables as follows: we defined K_b (or its increase N_1) to be the first exogenous variable and $K_b + K_r$ (or its increase N_2) to be the second exogenous noise, because then our elementary actions impact only one of the noise variables. Choosing K_r as the first exogenous term and $K_b + K_r$, for instance, would have our elementary actions let influence both.

The question where this notion of complexity of actions come from goes beyond the scope of this article. We hope that examples like Ex.1 showed that in real-life scenarios there are reasons to consider some actions as obviously more elementary than others. Further, we refer to the appendix where we argue that complexity of actions can be subject of scientific research and mention some approaches from modern physics.

Technically, the present article is also related to the work on causal discovery via interventions with unknown interventions targets [48] and joint causal inference (JCI) from multiple contexts [49], which both amount to assuming Markov properties and faithfulness in an augmented graph containing a "context variable" C [49] (also called "utility F-node" [48], "regime indicator variable" [17]) representing the action, change of context, or experimental conditions. We have avoided the assumption of faithfulness on purpose. While it can be helpful for causal discovery, we would not consider it fundamental enough to build a definition of causality on it, given that its validity can be questioned [34]. In a nutshell, the JCI approach of [49] requires faithfulness because it does not have the notion of *elementary* actions. Let us explain this for the toy example $C \to X_1 \to X_2$. Different contexts that change only $P(X_1)$ reveal the arrow $X_1 \to X_2$ *because of faithfulness*. Otherwise, we could assume $X_2 \to X_1$ and C influencing both such that $X_2 \perp\!\!\!\perp C \,|X_1$ holds. This can be excluded in our setting because our notion of elementary corresponds to a context variable that is connected to only one target node. This way, we do not need to add faithfulness or restrictions of model classes as in causal discovery. Note, for instance, that Definition 3 does not imply that actions in $\mathcal{A}_j$ influence *every* descendant of j in G (likewise for Definition 3). Accordingly, Definitions 3 and 3 will not specify the DAG uniquely.

6 Phenomenological causality couples to tangible causality

One can argue that a crucial property of causality is to describe the way a system with some variables couples to other variables in the world. In previous studies [50–52], causality is used to predict statistical relations of variables that have not been observed together. This section shows in which sense a causal DAG defined via phenomenological causality can be consistently embedded into the context of further variables. The following

15 Note that the difference between changing the noise and changing the functional relation between cause and effect disappears in the response function formulation of an FCM [47], where the noise is function valued and controls the deterministic map.

observations are fairly obvious and mostly known with respect to their purely mathematical content. Yet, we consider them crucial as justification of phenomenological causality.

6.1 Markov property of phenomenological causality

Let us first consider the mechanisms described by functions m_j in Definition 4. Since they represent structural equations $f_j(.,n_j)$ with fixed noise value n_j, we will denote them with superscript and write $m_j^{n_j}$. Whenever the noise values n_j are statistically independent across different statistical units, they induce a joint distribution that is Markovian with respect to G, see the study by Pearl [3], Theorem 1.4.1. We conclude that we obtain a Markovian joint distribution of $P(X_1, \ldots, X_n)$ whenever we control actions in $\mathcal{A}_j$ by independent random variables. The same holds true when we control the actions in Definition 3 by independent random variables and introduce formal random variables Θ_j controlling the causal conditionals $p^{\theta_j}(x_j|pa_j)$ (resembling the context variables in [49], for instance).

Then the joint distribution

$$\begin{aligned} P(X_1, \ldots, X_n) &= \int \prod_{j=1} p^{\theta_j}(x_j|pa_j) p(\theta_1) \cdots p(\theta_n) \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_n \\ &= \prod_{j=1}^{n} \int p^{\theta_j}(x_j|pa_j) p(\theta_j) \mathrm{d}\theta_j \end{aligned}$$

still factorizes with respect to G. Note that the assumption of independent Θ_j is in agreement with how [39] interpret the postulate of independent mechanisms (see the study by Zhang et al. [4], Section 2.1), namely, as *statistically independent changes* of the causal conditionals $p(x_j|pa_j)$ across environments. While Zhang et al. [39] use this property for the identification of the causal DAG, we use it to show that then the distribution averaging over different environments is still Markovian. In other words, G is true both with respect to each environment and with respect to the aggregated distribution. Moreover, for linear structural equations, e.g. the study by Bongers et al. (16), also the causal discovery method, LiNGAM would infer a causal structure that aligns with phenomenological causality, as mentioned earlier.

A more interesting scenario, however, is obtained when elementary actions are controlled by further random variables $Y_1, \ldots, Y_m$, which are connected by a non-trivial causal structure. We argue that then we obtain a joint distribution on $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ whose DAG is consistent with phenomenological causality. Assume, for instance, that some actions are not only controlled by independent noise variables N_j or Θ_j, respectively, but by one of the variables Y_i which are related by a DAG themselves. We then model the influence of Y_i on actions in $\mathcal{A}_j$ by introducing a second superscript to the mechanisms m_j and $p(x_j|pa_j)$, respectively, and obtain $m_j^{y_i,n_j}$ or $p^{y_i,\theta_j}(x_j|pa_j)$. Obviously, this way Y_i can be read as an additional parent of X_j. Further, let some Y_l be influenced by some X_j by modifying the structural equations for some Y_l such that they receive X_j as additional input.

We now define a directed graph with nodes $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ by drawing an edge from Y_i to X_j whenever Y_i controls actions in the set $\mathcal{A}_j$ and draw an edge from X_j to Y_l whenever the latter is influenced by the former. Whenever this graph is a DAG $\tilde{G}$, $P(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ will clearly be Markovian relative to $\tilde{G}$. This is because the generating process, by construction, follows structural equations according to $\tilde{G}$ and the joint distribution admits a corresponding Markov factorization.

In a scenario where causal relations among the Y_i and among Y_i and X_j are justified by tangible interventions, the abstract notion of causality between different X_j thus gets justified because it is consistent with the causal Markov condition also after embedding our abstract system into the tangible world. Getting back to our metaphor with a box with n knobs and n displays, our phenomenological definition of the causal relations inside the box is consistent with the DAG that describes causal relations between the box and the more tangible world, see Figure 12 for a visualization. Since the causal structures for our Examples 2, 7, and 8

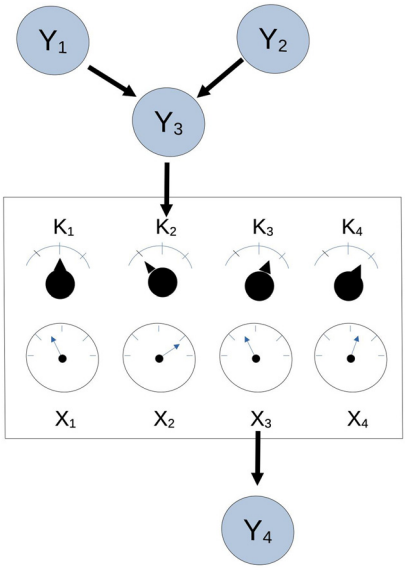

Figure 12: Visualization of a scenario where phenomenological causality couples to variables with tangible interventions. Then our construction of abstract causal relations between the variables X_j are justified by consistency in the sense of a Markov condition for the causal DAG of the joint system.

seemed particularly artificial, it arguably gets less artificial once such a system gets embedded into an environment with further variables.

6.2 Boundary consistency of the notion “elementary”

We emphasized that determining whether a variable X_j is “directly” affected by an action (and thus deciding whether the action is an intervention on X_j) is a causal question that may be equally hard to decide as the causal relations between the variables $X_1, \ldots, X_n$. Formally, the question can be phrased within a meta-DAG containing an additional variable A describing the actor, in which the relation between A and X_j appears as a usual link if and only if A is an intervention on X_j. Due to this arbitrariness of the boundary between system and actor, we expect a framework for causality to be consistent with respect to shifting this boundary (by extending or reducing the set of variables). In other words, descriptions need to be consistent regardless of where they draw the line between variables that are internal or external to the system [2, 49].[16] Likewise, our notion of “elementary” needs to be consistent with respect to boundary shifts; we show this for a certain class of marginalizations to subsets of variables. To explain this idea, we first introduce a rather strong notion of *causal sufficiency*:[17]

Definition 5. (Graphical causal sufficiency) Let $\mathbf{X} \coloneqq (X_1, \ldots, X_n)$ be nodes of a DAG G. A subset $\mathbf{X}_S$ is called graphically causally sufficient if there is no hidden common cause $C \notin \mathbf{X}_S$ that is causing at least two variables in $\mathbf{X}_S$ (and the causing paths go only through nodes that are not in $\mathbf{X}_S$).

16 In the early days of quantum physics, Heisenberg described a similar consistency of the theory with respect to shifting the boundary between *measurement apparatus* and *quantum system to be measured* (the “Heisenberg cut”). Reference [53] is even more similar in spirit to our boundary consistency because it describes the arbitrariness of the boundary between *controlling device* and *system to be controlled* for interventions on microscopic physical systems and constructs a framework for physical controllers in which this boundary can be shifted in a consistent way.

17 Or just called ‘causal suffciency’ as in [4], but there the sentence in the bracket has been forgotten, as noted in the errata of the book.

In general, the model class of causal DAGs is not closed under marginalization, but requires other model classes, such as maximal ancestral graphs [54] or acyclic directed mixed graphs [55]. Here, we restrict the attention to the simple case of graphical causal sufficiency, where the causal model remains in the class of DAGs after marginalization:

Definition 6. (Marginal DAG) Let $\mathbf{X}$ be the nodes of a DAG G and $\mathbf{X}_S$ a graphically causally sufficient set. Then the marginalization G_S of G to the nodes $\mathbf{X}_S$ is the DAG with nodes $\mathbf{X}_S$ and an edge $X_i \to X_j$ whenever there exists a directed path from X_i to X_j in G containing no node from $\mathbf{X}_S$ (except X_i, X_j).

To justify the definition, we need to show that the distribution of $\mathbf{X}_S$ is Markov relative to G_S and that G_S correctly describes interventional probabilities. It is easy to check the Markov condition: Let $\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C$ be subsets of $\mathbf{X}_S$ such that $\mathbf{X}_A$ is d-separated from $\mathbf{X}_B$ by $\mathbf{X}_C$ in G, hence every path in G connecting a node in $\mathbf{X}_A$ with one in $\mathbf{X}_B$ contains either (i) a chain or a fork with middle node in $\mathbf{X}_C$ or (ii) an inverted fork whose middle node is not $\mathbf{X}_C$ and also not its descendants. It is easy to see that conditions (i) and (ii) are preserved when directed paths are collapsed to single arrows, and thus, the same conditions hold in G_S. To see that interventions on arbitrary nodes in $\mathbf{X}_S$ can equivalently be computed from G_S, we recall that interventional probabilities can be computed from backdoor adjustments [3] (Equation (3.19)). We can easily verify that if $Z \subset \mathbf{X}_S$ satisfies the backdoor criterion in G_S relative to an ordered pair (X_i, X_j) of variables in $\mathbf{X}_S$, it also satisfies it in G because the property of blocking backdoor paths is inherited from G.

The following result shows that our notion of "elementary" is preserved under marginalization to causally sufficient subsets:

Theorem 1. (boundary consistency) *Let G be a DAG with nodes $\mathbf{X} := \{X_1, \ldots, X_n\}$ and $P, \tilde{P}$ be joint distributions of $\mathbf{X}$ that are Markov relative to G and differ only by one term in the factorization* (1). *For some subset S of nodes satisfying graphical causal sufficiency, let G_S with $\mathbf{X}_S \subset \mathbf{X}$ be a marginalization of G, and $P_S, \tilde{P}_S$ be marginalizations of $P, \tilde{P}$, respectively. Then P_S and $\tilde{P}_S$ also differ by one conditional at most.*

Proof. Let P and $\tilde{P}$ differ by the conditional corresponding to X_j. Introduce a binary variable I pointing on X_j which controls switching between $P(X_j|PA_j)$ and $\tilde{P}(X_j|PA_j)$. Formally, we thus define a distribution $\hat{P}$ on $(\mathbf{X}, I)$ such that $\hat{P}(x_j|pa_j, I = 0) = P(x_j|pa_j)$ and $\hat{P}(x_j|pa_j, I = 1) = \tilde{P}(x_j|pa_j)$. Let G^I be the augmented DAG containing the nodes of G and I with an arrow from I to X_j. For the case where $X_j \in \mathbf{X}_S$, it is sufficient to show that the marginalization of G^I to $S \cup \{I\}$ does not connect I with any node X_i other than X_j, which follows already from the fact that any directed path from I to X_i passes X_j. Now assume that X_j is not in $\mathbf{X}_S$. By causal sufficiency of $\mathbf{X}_S$, there is a unique node $X_{\tilde{j}}$ among the descendant of $X_j \in \mathbf{X}_S$ that blocks all paths to other nodes in $\mathbf{X}_S$ (otherwise $\mathbf{X}_S$ would not be causally sufficient). Hence, the DAG G^I_S contains only an edge to $X_{\tilde{j}}$ but no other node in $\mathbf{X}_S$ (Figure 13).

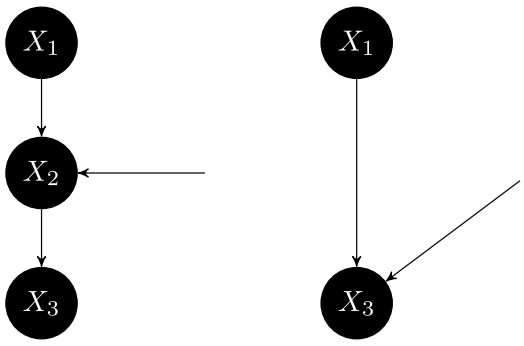

Figure 13: Action on node X_2 (left), which results in an action on X_3 after dropping node X_2 (right).

It seems that every framework that is supposed to be general enough to describe significant aspects of the world should not only be able to describe the system under consideration but also its interaction with agents. Understanding why and in what sense certain actions are more elementary than others is still a question to be answered outside the framework. However, demanding consistency of different boundaries between system and intervening agents seems a more modest and feasible version of "understanding" of how to define elementary. After all, it is the description of the system that decides which variables are internal or external, as also noted in Pearl's Epilogue [56].

7 Conclusions

We have described several scenarios – some of them are admittedly artificial, but some of them are closer to real-life problems – where causal relation between observed quantities are not defined *a priori*, but get only well-defined after specifying the "elementary actions" that are considered interventions on the respective variables. We have argued that this specification admits the definition of an abstract notion of causality in domains where the mechanistic view of tangible causal interactions fails. We believe that this approach renders the context dependence of causality more transparent since there may be different *elementary actions* in different contexts. It is possible that at least some part of the fuzziness of some relevant causal questions (e.g. "does income influence life expectancy?") comes from the missing specification of actions. From this point of view, one could argue to accept only causal questions that directly refer to the treatment effect for which *the treatment itself* is obviously a feasible action (e.g. taking a drug or not) and rejecting questions about the causal effect of variables like "income." However, our approach is different in the sense that – after having defined the elementary actions – it does talk about causal relations between variables "inside the box of abstract variables," that is, variables for which interventions are not defined *a priori*. This is because we believe that analyzing causal relations "inside the box" is crucial for understanding complex systems.

Acknowledgements: Many thanks to Joris Mooij for inspiring discussions on the relation to [20].

Conflict of interest: The authors declare no competing interests.

References

[1] Pearl J, Mackenzie J. The book of why. USA: Basic Books; 2018.
[2] Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. New York, NY: Springer-Verlag; 1993.
[3] Pearl J. Causality. Cambridge: Cambridge University Press; 2000.
[4] Peters J, Janzing D, Schölkopf B. Elements of causal inference - foundations and learning algorithms. Cambridge, MA: MIT Press; 2017.
[5] Zhang A, Lipton ZC, Li M, Smola AJ. Dive into deep learning. 2020. https://d2l.ai.
[6] Database with cause–effect pairs. https://webdav.tuebingen.mpg.de/cause–effect/. Copyright information for each cause–effect pair is contained in the respective description file.
[7] Guyon I, Statnikov A, Bakir-Batu B. Cause effect pairs in machine learning. The Springer Series on Challenges in Machine Learning Berlin & Heidelberg: Springer; 2019 Jan.
[8] Ke NR, Wang JX, Mitrovic J, Szummer M, Rezende DJ. Amortized learning of neural causal representations, 2020. arxiv:2008.09310.
[9] Hernán MA, Taubman SL. Does obesity shorten life? the importance of well-defined interventions to answer causal questions. Int J Obes. 2008;32:8–14.
[10] Lauritzen S. Graphical models. Statistical Science Series edition. Oxford, New York, Oxford: Clarendon Press; 1996.
[11] Janzing D, Blöbaum P, Minorics L, Faller P. Quantifying causal contribution via structure preserving interventions. 2020.
[12] Chalupka K, Perona P, Eberhardt F. Visual causal feature learning. In Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence, UAI'15. AUAI Press; 2015. p. 181–90.

[13] Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, Goyal A, et al. Towards causal representation learning. Proc IEEE. 2021;109(5):1–23.
[14] Woodward J. Making things happen. New York, NY: Oxford University Press; 2003.
[15] Baumgartner M. Interdefining causation and intervention. Dialectica. 2009;63(2):175–94.
[16] Bongers S, Blom T, Mooij JM. Causal modeling of dynamical systems. 2018. arXiv:1803.08784.
[17] Dawid P. Decision-theoretic foundations for statistical causality. J Causal Inference. 2021;9(1):39–77.
[18] Hausman DM. Causal asymmetries. Cambridge Studies in Probability, Induction and Decision Theory. Cambridge: Cambridge University Press; 1998.
[19] Reichenbach H. The direction of time. Berkeley: University of California Press; 1956.
[20] Blom T, Van Diepen MM, Mooij JM. Conditional independences and causal relations implied by sets of equations. J Mach Learn Res. Jan 2021;22(1):8044–105.
[21] Goldstein H, Poole C, Safko J. Classical mechanics. 3rd edition. Boston, MA: Pearson; 2002.
[22] Mooij J, Janzing D, Schölkopf B. From ordinary differential equations to structural causal models: the deterministic case. In: Nicholson A, Smyth P, editors. Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI). Oregon, USA: AUAI Press Corvallis; 2013. p. 440–8.
[23] Blom T, Bongers S, Mooij JM. Beyond structural causal models: Causal constraints models. In: Globerson A, Silva R, editors. Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, July 22–25, 2019, vol. 115 of Proceedings of Machine Learning Research. Tel Aviv, Israel: AUAI Press; 2019. p. 585–94.
[24] Lauritzen SL, Richardson TS. Chain graph models and their causal interpretations. J R Stat Soc Ser B-Stat Methodol. 2002;64(3):321–48.
[25] Einstein A. Relativity: the special and general theory. New York: H. Holt and Company; 1920.
[26] Hoyer P, Shimizu S, Kerminen A, Palviainen M. Estimation of causal effects using linear non-gaussian causal models with hidden variables. Int J Approx Reason. 2008;49(2):362–78.
[27] Janzing D, Peters J, Mooij J, Schölkopf B. Identifying latent confounders using additive noise models. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI 2009). Ng A, Bilmes J, editors. Corvallis, OR, USA: AUAI Press; 2009. p. 249–57.
[28] Bowden R, Turkington D. Instrumental variables. Cambridge: Cambridge University Press; 1984.
[29] Mastakouri A, Schölkopf B, Janzing D. Selecting causal brain features with a single conditional independence test per feature. In: Wallach H, Larochelle H, Beygelzimer A, d Alché-Buc F, Fox E, Garnett R, editors. Advances in neural information processing systems. Vol. 32. Red Hook, NY: Curran Associates, Inc.; 2019. p. 1–12.
[30] Spirtes PL, Scheines R. Causal inference of ambiguous manipulations. Philosophy Sci. 2004;71:833–45.
[31] Rubenstein PK, Weichwald S, Bongers S, Mooij JM, Janzing D, Grosse-Wentrup M, et al. Causal consistency of structural equation models. In: Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence (UAI 2017). 2017.
[32] Beckers S, Halpern JY. Abstracting causal models. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019. AAAI Press; 2019. p. 2678–85.
[33] Janzing D, Schölkopf B. Causal inference using the algorithmic Markov condition. IEEE Trans Inform Theory. 2010;56(10):5168–94.
[34] Lemeire J, Janzing D. Replacing causal faithfulness with algorithmic independence of conditionals. Minds Machines. 2012;23(2):227–49.
[35] Schölkopf B, Janzing D, Peters J, Sgouritsa E, Zhang K, Mooij J. On causal and anticausal learning. In: Langford J, Pineau J, editors. Proceedings of the 29th International Conference on Machine Learning (ICML). ACM; 2012. p. 1255–62.
[36] Sugiyama M, Kawanabe M. Machine learning in non-stationary environments: introduction to covariate shift adaptation. Cambridge, MA: The MIT Press; 2012.
[37] Lewis D. Counterfactual dependence and timeas arrow. Noûs. 1979;13(4):455–76.
[38] Simon H. Studies in Econometric Methods, chapter Causal ordering and identifiability. Hoboken, NJ: John Wiley & Sons; 1953. p. 49–74.
[39] Zhang K, Huang B, Zhang J, Glymour C, Schölkopf B. Causal discovery from nonstationary/heterogeneous data: Skeleton estimation and orientation determination. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17. 2017. p. 1347–53.
[40] Zang Y, Zou Q, Fu T, Ng F, Fowler B, Yang J, et al. Directing isomerization reactions of cumulenes with electric fields. Nature 2019;10(1):4482.
[41] Kano Y, Shimizu S. Causal inference using nonnormality. In Proceedings of the International Symposium on Science of Modeling, the 30th Anniversary of the Information Criterion, Tokyo, Japan; 2003. p. 261–70.
[42] Dash D. Restructuring dynamic causal systems in equilibrium. In: Cowell RG, Ghahramani Z, editors, Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, volume R5 of Proceedings of Machine Learning Research, PMLR; 06-08 Jan 2005. p. 81–8. Reissued by PMLR on 30 March 2021.
[43] Druzdzel MJ, Van Leijen H. Causal reversibility in Bayesian networks. J Experiment Theoretic Artif Intel. 2001;13(1):45–62.
[44] Zhang J, Spirtes P. Intervention, determinism, and the causal minimality condition. Synthese. 2011;182:335–47.
[45] Moneta A, Entner D, Hoyer P, Coad A. Causal inference by independent component analysis: Theory and applications. Oxford Bullet Econ Stat. 2013;75(5):705–30.

[46] Hausman DM, Woodward J. Independence, invariance and the causal Markov condition. British Soc Philosophy Sci. 1999;50:521–83.
[47] Balke A, Pearl J. Counterfactual probabilities: Computational methods, bounds, and applications. In: Lopez R Mantaras D, Poole D, editors. Uncertainty in Artifical Intelligence. vol. 10. San Mateo: Morgan Kaufmann; 1994.
[48] Jaber A, Kocaoglu M, Shanmugam K, Bareinboim E. Causal discovery from soft interventions with unknown targets: Characterization and learning. In Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in Neural Information Processing Systems. Vol. 33. Curran Associates, Inc.; 2020. p. 9551–61.
[49] Mooij JM, Magliacane S, Claassen T. Joint causal inference from multiple contexts. J Mach Learn Res. Jan 2020;21(1):1–108.
[50] Tsamardinos I, Triantafillou S, Lagani V. Towards integrative causal analysis of heterogeneous data sets and studies. J Mach Learn Res. 2012;13(1):1097–157.
[51] Janzing D. Merging joint distributions via causal model classes with low vc dimension. 2018. arXiv: http://arXiv.org/abs/arXiv:1804.03206.
[52] Gresele L, Von Kügelgen J, Kübler J, Kirschbaum E, Schölkopf B, Janzing D. Causal inference through the structural causal marginal problem. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S, editors. Proceedings of the 39th International Conference on Machine Learning, volume 162 of Proceedings of Research. PMLR; 17–23 Jul 2022. p. 7793–824.
[53] Janzing D. Is there a physically universal cellular automaton or Hamiltonian?. 1720. http://arXiv.org/abs/arXiv:1009.
[54] Richardson T, Spirtes P. Ancestral graph Markov models. Ann Stat. 2002;30(4):962–1030.
[55] Bongers S, Forré P, Peters J, Mooij JM. Foundations of structural causal models with cycles and latent variables. Ann Stat. 2021;49(5):2885–915.
[56] Pearl J. The art and science of cause and effect. 2nd edition, Cambridge, UK: Cambridge University Press; 2009. p. 401–28.
[57] Papadimitriou CH. Computational complexity. Hoboken, NJ: John Wiley and Sons Ltd., GBR; 2003. p. 260–5.
[58] Nielsen M, Chuang I. Quantum computation and quantum information. Cambridge: Cambridge University Press; 2000.
[59] Deutsch D. The fabric of reality. London, UK: The Penguin Press; 1997.
[60] Fernandez J, Lloyd S, Mor T, Roychowdhury V. Algorithmic cooling of spins: A practicable method for increasing polarization. Int J Quant Inf. 2004;2(4):461–7.
[61] Janzing D. On the computational power of molecular heat engines. J Stat Phys. 2006;122(3):531–56.
[62] Wocjan P, Janzing D, Decker T. Measuring 4-local n-qubit observables could probabilistically solve PSPACE. Quantum Inform Comput. 2008;4(8 and 9):741–55.
[63] Yosi A, Aharonov D. Fast-forwarding of hamiltonians and exponentially precise measurements. Nature Commun. 2017;8:11.
[64] Janzing D. Computer science approach to quantum control. Habilitationsschrift: UniVerlag Karlsruhe; 2006.

Appendix

A Complexity of actions in modern micro-physics

Notions of complexity of transformations have traditionally been subject of computer science in the sense of *computational* complexity. In a nutshell, computational complexity explores how the number of *elementary logical operations* scales with the problem size. While the complexity theory does not come with an advice which logical transformations are supposed to be elementary, the asymptotic scaling behaviour is independent of this convention provided that they can be defined as operations of a universal Turing machine [57]. Computer science has therefore considered different models of computation as the basis for complexity theory. While asymptotic behaviour can be a good heuristic to estimate running time for real problems, the question where a notion of complexity in our finite world should come from remains actually open.

However, there are ideas from modern physics that provide new insights in this regard. The preceding three decades of quantum information research [58] has intertwined computer science and physics in a way that the disciplines have never seen before. First, Deutsch [59] emphasized that *the laws of physics* determines which logical operations are simple and complex and argued that the laws of quantum physics entail a new notion of *Quantum Complexity* – which may differ from complexity of classical computer science – but seems more fundamental since it is a notion of complexity that is defined via microscopic physical processes. There, logical operations are considered elementary because one can describe existing physical interactions that implement them. Second, the elementary operations in quantum computing [58] can not only be interpreted as *logical* operations, but also as operations whose goal is more general than only implementing a computation. Researchers considered, for instance, the complexity of *cooling algorithms*, that is, complex transformations on molecular systems that transfer heat from one part to the other [60], and similarly, the complexity of *heat engines* [61]. Further, several articles considered the complexity of measurement processes [62,63] and compared their complexity with hard *computational* tasks (see also the study by Janzing [64] for a slightly outdated overview). The essential message for the present article is that complexity of actions is indeed despite the fuzziness of the question, subject of the scientific research.