Causal Consistency of Structural Equation Models

Paul K. Rubenstein[*12], **Sebastian Weichwald**[*13], **Stephan Bongers**[4], **Joris M. Mooij**[4]
Dominik Janzing[1], **Moritz Grosse-Wentrup**[1], **Bernhard Schölkopf**[1]
[*]Equal contribution
[1]Empirical Inference, MPI for Intelligent Systems, [2]Machine Learning Group, University of Cambridge,
[3]Max Planck ETH Center for Learning Systems, [4]Informatics Institute, University of Amsterdam

Abstract

Complex systems can be modelled at various levels of detail. Ideally, causal models of the same system should be consistent with one another in the sense that they agree in their predictions of the effects of interventions. We formalise this notion of consistency in the case of Structural Equation Models (SEMs) by introducing *exact transformations* between SEMs. This provides a general language to consider, for instance, the different levels of description in the following three scenarios: (a) models with large numbers of variables versus models in which the 'irrelevant' or unobservable variables have been marginalised out; (b) micro-level models versus macro-level models in which the macro-variables are aggregate features of the micro-variables; (c) dynamical time series models versus models of their stationary behaviour. Our analysis stresses the importance of well specified interventions in the causal modelling process and sheds light on the interpretation of cyclic SEMs.

1 INTRODUCTION

Physical systems or processes in the real world are complex and can be understood at various levels of detail. For instance, a gas in a volume consists of a large number of molecules. But instead of modelling the motions of each particle individually (micro-level), we may choose to consider macroscopic properties of their motions such as temperature and pressure. Our decision to use such macroscopic properties is first necessitated by practical considerations. Indeed, for all but extremely simple cases, making a measurement of all the individual molecules is practically impossible and our resources insufficient for modelling the $\sim 10^{22}$ particles present per litre of ideal gas.

Furthermore, the decision for a macroscopic description level is also a pragmatic one: if we only wish to reason about temperature and pressure, a model of 10^{22} particles is ill-suited.

Statistical physics explains how higher-level concepts such as temperature and pressure arise as statistical properties of a system of a large number of particles, justifying the use of a macro-level model as a useful transformation of the micro-level model [1]. However, in many cases aggregate or indirect measurements of a complex system form the basis of a macroscopic description of the system, with little theory to explain whether this is justified or how the micro- and macro-descriptions stand in relation to each other.

Due to deliberate modelling choice or the limited ability to observe a system, differing levels of model descriptions are ubiquitous and occur, amongst possibly others, in the following three settings:

(a) Models with large numbers of variables versus models in which the 'irrelevant' or unobservable variables have been marginalised out [3]; e. g. modelling blood cholesterol levels and risk of heart disease while ignoring other blood chemicals or external factors such as stress.

(b) Micro-level models versus macro-level models in which the macro-variables are aggregate features of the micro-variables [4, 5, 9, 11, 17]; e. g. instead of modelling the brain as consisting of 100 billion neurons it can be modelled as averaged neuronal activity in distinct functional brain regions.

(c) Dynamical time series models versus models of their stationary behaviour [6, 8, 11–13, 15]; e. g. modelling only the final ratios of reactants and products of a time evolving chemical reaction.

In the context of causal modelling, such differing model levels should be consistent with one another in the sense that they agree in their predictions of the effects of interventions. The particular causal models we focus on in this
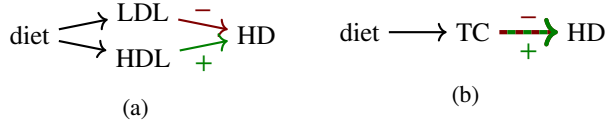
Figure 1: As illustrated by (a), the current consensus is that LDL (resp. HDL) has a negative (resp. positive) effect on heart disease (HD). Considering TC = LDL + HDL to be a causal variable as in (b) leads to problems: two diets promoting raised LDL levels and raised HDL levels have the same effect on TC but opposite effects on heart disease. Hence different studies come to contradictory conclusions about the effect of TC on heart disease.

paper are Structural Equation Models (SEMs, Section 2, Section 3) [16, 19].

In Section 4, we introduce the notion of an exact transformation between two SEMs, providing us with a general framework to evaluate when two models can be thought of as causal descriptions of the same system. An important novel idea of this paper is to explicitly make use of a natural ordering on the set of interventions. On a high level, if an SEM can be viewed as an exact transformation of another SEM, we are provided with an explicit correspondence between the two models in such a way that causal reasoning on both levels is consistent. We discuss this notion of consistency in detail in Sections 4.4 and 4.5.

In Section 5 we apply this mathematical framework and prove the exactness of transformations belonging to each of the three categories listed above, with practical implications for the following questions in causal modelling: When can we model only a subsystem of a more complex system? When does a micro-level system admit a causal description in terms of macro-level features? How do cyclic SEMs arise? The fact that these distinct problems can all be considered using the language of transformations between SEMs demonstrates the generality of our approach. We close in Section 6 with a discussion.

1.1 A HISTORICAL MOTIVATION: CHOLESTEROL AND HEART DISEASE

In the following we give an example of the problems that can arise when there exists no consistent correspondence between two causal models, i. e. neither model can be viewed as an exact transformation of the other. This example falls into category (b) of the differing model levels listed above and was used by [18] to illustrate problems in the causal modelling process.

Historically, the level of total cholesterol in the blood (TC) was thought to be an important variable in determining risk of heart disease (HD). To investigate this, differ-

ent experiments were carried out in which patients were assigned to different diets in order to raise or lower TC. Conflicting evidence was found by different experiments: some found that higher TC had the effect of lowering HD, while others found the opposite (cf. Figure 1b) [20, 21].

From our point of view, this problem (seemingly conflicting studies) arose from trying to perform an 'invalid' transformation of the 'true' underlying model (cf. Figure 1a). According to the American Heart Association, the current scientific consensus is that the two types of blood cholesterol, low-density lipoprotein (LDL) and high-density lipoprotein (HDL), have a negative and positive effect on HD respectively. Assigning diets that raise LDL or HDL both raise TC but have different effects on HD. It is therefore not possible to transform the model in Figure 1a into the model in Figure 1b without leading to conflict: in order to reason about the causes of HD we need to consider the variables LDL and HDL separately.

2 STRUCTURAL EQUATION MODELS

SEMs are a widely used framework in causal modelling, with applications in neuroscience, economics and the social sciences [2, 16]. In this section we introduce them as an abstract mathematical object; in Section 3 we describe their use as a causal modelling tool. Readers already familiar with SEMs should note that our definition is more general and deviates from the standard definition of SEMs in the following ways: we do not require that all possible perfect interventions be modelled; we do not assume independence of exogenous variables;[1] and we do not require acyclicity.

Definition 1 (Structural Equation Model (SEM))**.** *Let $\mathbb{I}_X$ be an index set. An SEM $\mathcal{M}_X$ over variables $X = (X_i : i \in \mathbb{I}_X)$ taking value in $\mathcal{X}$ is a triple $(\mathcal{S}_X, \mathcal{I}_X, \mathbb{P}_E)$ where*

- *$\mathcal{S}_X$ is a set of structural equations, i. e. it is a set of equations $X_i = f_i(X, E_i)$ for $i \in \mathbb{I}_X$;*
- *$(\mathcal{I}_X, \leq_X)$ is a subset of all perfect interventions equipped with a natural partial ordering (see below), i. e. it is an index set where each index corresponds to a particular perfect intervention on some of the X variables;*
- *$\mathbb{P}_E$ is a distribution over the exogenous variables $E = (E_i : i \in \mathbb{I}_X)$;*
- *with $\mathbb{P}_E$-probability one, under any intervention $i \in \mathcal{I}_X$ there is a unique solution $x \in \mathcal{X}$ to the intervened structural equations. This ensures that for any*

---

[1]Exogenous variables are also referred to as *noise variables* in the literature. Our relaxation of the assumption of independent exogenous variables means our models may be considered a type of semi-Markovian causal model.

intervention $i \in \mathcal{I}_X$, $\mathcal{M}_X$ induces a well-defined distribution over $\mathcal{X}$.[2]

In an SEM, each X_i is a function of the X-variables and the exogenous variable E_i. In this mathematical model, a perfect intervention on a single variable $\mathrm{do}(X_i = x_i)$ is realised by replacing the structural equation for variable X_i in S_X with $X_i = x_i$. Perfect interventions on multiple variables, e.g. $\mathrm{do}(X_i = x_i, X_j = x_j)$, are similarly realised by replacing the structural equations for each variable individually. Elements of $\mathcal{I}_X$ correspond to perfectly intervening on a subset of the X variables, setting them to some particular combination of values.

$\mathcal{I}_X$ has a natural partial ordering in which, for interventions $i, j \in \mathcal{I}_X$, $i \leq_X j$ if and only if i intervenes on a subset of the variables that j intervenes on and sets them equal to the same values as j. For example, $\mathrm{do}(X_i = x_i) \leq_X \mathrm{do}(X_i = x_i, X_j = x_j)$.[3] The observation that this structure is important is a contribution of this paper. We make crucial use of it in the next section.

The purpose of the following example is to illustrate how SEMs are written in our notation and to provide and example of a restricted set of interventions $\mathcal{I}_X$.

Example 2. *Consider the following SEM defined over the variables $\{B_1, B_2, L\}$*

$$S_X = \left\{ B_1 = E_1, \ B_2 = E_2, \ L = \mathrm{OR}(B_1, B_2, E_3) \right\}$$
$$\mathcal{I}_X = \left\{ \varnothing, \ \mathrm{do}(B_1 = 0), \ \mathrm{do}(B_2 = 0), \right.$$
$$\left. \mathrm{do}(B_1 = 0, B_2 = 0) \right\},$$

$$\{E_1, E_2, E_3\} \overset{iid}{\sim} \mathrm{Bernoulli}(0.5)$$

where by the element $\varnothing \in \mathcal{I}$ we denote the null-intervention corresponding to the unintervened SEM.

3   SEMS FOR CAUSAL MODELLING

In addition to being abstract mathematical objects, SEMs are used in causal modelling to describe distributions of variables and how they change under interventions [16]. The do-interventions as abstract manipulations of SEMs are understood as corresponding to actual (or potentially

---

[2]That is, with probability one over the exogenous variables E, for each draw $E = e$ there exists a unique value $x \in \mathcal{X}$ such that e and x satisfy the intervened structural equations. The distribution of E in conjunction with S_X then implies a distribution over $\mathcal{X}$ for each intervention $i \in \mathcal{I}_X$ via these unique solutions. If the SEM is acyclic, this is always satisfied; we impose this condition because we also consider *cyclic* SEMs [3].

[3]Informally, this means that j can be performed after i without having to change or undo any of the changes to the structural equations made by i. Not all pairs of elements must be comparable: for instance, if $i = \mathrm{do}(X_1 = x_1)$ and $j = \mathrm{do}(X_2 = x_2)$, then neither $i \leq_X j$ nor $j \leq_X i$.

only hypothetical) physical implementations in the real world, i. e. the model is 'rooted in reality'. For instance, if a binary variable B_1 in an SEM reflects whether a light bulb is emitting light, then $\mathrm{do}(B_1 = 0)$ could be achieved by flipping the light switch or by removing the light bulb.

The SEM in Example 2 could be thought of as a simple causal model of two light bulbs B_1 and B_2 and the presence of light L in a room with a window. Suppose that we have no access to the light switch and there are no curtains in the room but that we can intervene by removing the light bulbs. We can model this restricted set of interventions by $\mathcal{I}_X$, i. e. the do-intervention on the SEM side $\mathrm{do}(B_1 = 0)$ corresponds to removing the light bulb B_1.

The partial ordering of $\mathcal{I}_X$ corresponds to the ability to compose physical implementations of interventions. The fact that we can first remove light bulb B_1 ($\mathrm{do}(B_1 = 0)$) and then afterwards remove light bulb B_2 (resulting in the combined intervention $\mathrm{do}(B_1 = 0, B_2 = 0)$) is reflected in the partial ordering via the relation $\mathrm{do}(B_1 = 0) \leq_X \mathrm{do}(B_1 = 0, B_2 = 0)$.

4   TRANSFORMATIONS BETWEEN SEMS

We now work towards our definition of an exact transformation between SEMs. Our core idea is to analyse the correspondence between different levels of modelling by considering one model to be a transformation of the other. We discuss in Section 4.4 how causal reasoning in two SEMs relate when one SEM can be viewed as an exact transformation of the other and in Section 4.5 we illustrate what can go wrong when this is not the case.

4.1   DISTRIBUTIONS IMPLIED BY AN SEM

Usually, a statistical model implies a single joint distribution over all variables once its parameters are fixed. SEMs are different in that, once the parameters are fixed, an SEM implies a family of joint distributions over the random variables, one for each intervention. That is, for each intervention $i \in \mathcal{I}_X$, the SEM $\mathcal{M}_X$ defines a distribution over $\mathcal{X}$ which we denote by $\mathbb{P}_X^{\mathrm{do}(i)}$. Throughout, we will denote the null-intervention corresponding to the unintervened setting by $\varnothing \in \mathcal{I}_X$. We can write the poset of all distributions implied by the SEM $\mathcal{M}_X$ as

$$\mathcal{P}_X := \left(\left\{ \mathbb{P}_X^{\mathrm{do}(i)} \ : \ i \in \mathcal{I}_X \right\}, \leq_X \right)$$

where $\leq_X$ is the partial ordering inherited from $\mathcal{I}_X$, i. e. $\mathbb{P}_X^{\mathrm{do}(i)} \leq_X \mathbb{P}_X^{\mathrm{do}(j)} \iff i \leq_X j$.[4]

---

[4]More formally, one would need to define $\mathcal{P}_X$ to be the poset of *tuples* $\left(i, \mathbb{P}_X^{\mathrm{do}(i)} \right)$ to avoid problems in the case that $\mathbb{P}_X^{\mathrm{do}(i)} =$

Note that $\mathcal{P}_X$ contains all of the information in $\mathcal{M}_X$ about the different distributions implied by the SEM and, importantly, how they are related via the interventions.[5]

4.2 TRANSFORMATIONS OF RANDOM VARIABLES

Suppose we have a function $\tau : \mathcal{X} \to \mathcal{Y}$ which maps the variables of the SEM $\mathcal{M}_X$ to another space $\mathcal{Y}$. Observe that since X is a random variable, $\tau(X)$ is also a random variable. For any distribution $\mathbb{P}_X$ on $\mathcal{X}$ we thus obtain the distribution of the variable $\tau(X)$ on $\mathcal{Y}$ as $\mathbb{P}_{\tau(X)} = \tau\left(\mathbb{P}_X\right)$ via the push-forward measure.

In particular, for each intervention $i \in \mathcal{I}_X$ we can define the induced distribution $\mathbb{P}^i_{\tau(X)} = \tau\left(\mathbb{P}^{\mathrm{do}(i)}_X\right)$. We can write the poset of distributions on $\mathcal{Y}$ that are induced by the original SEM $\mathcal{M}_X$ and the transformation τ as

$$\mathcal{P}_{\tau(X)} := \left(\left\{\mathbb{P}^i_{\tau(X)} \;:\; i \in \mathcal{I}_X\right\}, \leq_X\right)$$

where $\leq_X$ is the partial ordering inherited from $\mathcal{P}_X$ (and in turn from $\mathcal{I}_X$).

$\mathcal{P}_{\tau(X)}$ is just a structured collection of distributions over $\mathcal{Y}$, indexed by interventions $\mathcal{I}_X$ on the $\mathcal{X}$-level; importantly, the indices are *not* interventions on the $\mathcal{Y}$-level.

4.3 EXACT TRANSFORMATIONS BETWEEN SEMS

Although $\mathcal{P}_{\tau(X)}$ is a poset of distributions over $\mathcal{Y}$, there does not necessarily exist an SEM $\mathcal{M}_Y$ over $\mathcal{Y}$ that implies it. For instance, if there is some intervention $i \in \mathcal{I}_X \setminus \{\varnothing\}$ such that none of the variables Y_i is constant under the distribution $\mathbb{P}^i_{\tau(X)}$, then $\mathbb{P}^i_{\tau(X)}$ could not possibly be expressed as arising from a do-intervention $j \in \mathcal{I}_Y \setminus \{\varnothing\}$ in any SEM over $\mathcal{Y}$.[6]

The case in which there *does* exist an SEM $\mathcal{M}_Y$ that implies $\mathcal{P}_{\tau(X)}$ is special, motivating our main definition.

Definition 3 (Exact Transformations between SEMs)**.** *Let $\mathcal{M}_X$ and $\mathcal{M}_Y$ be SEMs and $\tau : \mathcal{X} \to \mathcal{Y}$ be a function. We say $\mathcal{M}_Y$ is an* exact τ-transformation *of $\mathcal{M}_X$ if there exists a surjective order-preserving map $\omega : \mathcal{I}_X \to \mathcal{I}_Y$ such that*

$$\mathbb{P}^i_{\tau(X)} = \mathbb{P}^{\mathrm{do}(\omega(i))}_Y \quad \forall i \in \mathcal{I}_X$$

where $\mathbb{P}^i_{\tau(X)}$ *is the distribution of the $\mathcal{Y}$-valued random variable $\tau(X)$ with $X \sim \mathbb{P}^{\mathrm{do}(i)}_X$.*

Order-preserving means that $i \leq_X j \implies \omega(i) \leq_Y \omega(j)$. It is important that the converse need not in general hold as this would imply that ω is injective,[7] and hence also bijective. This would constrain the ways in which $\mathcal{M}_Y$ can be 'simpler' than $\mathcal{M}_X$.[8] That ω is surjective ensures that for any do-intervention $j \in \mathcal{I}_Y$ on $\mathcal{M}_Y$ there is at least one corresponding intervention on the $\mathcal{M}_X$ level, namely an element of $\omega^{-1}(\{j\}) \subseteq \mathcal{I}_X$. The following two results follow immediately from the definition (cf. proofs in Appendix A).

Lemma 4. *The identity mapping and permuting the labels of variables are both exact transformations.*

This is a good sanity check; it would be problematic if this were not the case and the labelling of our variables mattered. Similarly, compositions of exact transformations are also exact.

Lemma 5 (Transitivity of exact transformations)**.** *If $\mathcal{M}_Z$ is an exact τ_{ZY}-transformation of $\mathcal{M}_Y$ and $\mathcal{M}_Y$ is an exact τ_{YX}-transformation of $\mathcal{M}_X$, then $\mathcal{M}_Z$ is an exact $(\tau_{ZY} \circ \tau_{YX})$-transformation of $\mathcal{M}_X$.*

The following theorem is a consequence of the fact that ω is order-preserving. This is a mathematical formalisation of the sense in which an exact transformation preserves causal reasoning, which will be elaborated upon in the next subsection.

Theorem 6 (Causal consistency under exact transformations)**.** *Suppose that $\mathcal{M}_Y$ is an exact τ-transformation of $\mathcal{M}_X$ and ω is a corresponding surjective order-preserving mapping between interventions. Let $i, j \in \mathcal{I}_X$ be interventions such that $i \leq_X j$. Then the following diagram commutes:*

$$
\begin{array}{ccccc}
\mathbb{P}_X & \xrightarrow{\mathrm{do}(i)} & \mathbb{P}^{\mathrm{do}(i)}_X & \xrightarrow{\mathrm{do}(j)} & \mathbb{P}^{\mathrm{do}(j)}_X \\
\downarrow{\scriptstyle\tau} & & \downarrow{\scriptstyle\tau} & & \downarrow{\scriptstyle\tau} \\
\mathbb{P}_Y & \xrightarrow{\mathrm{do}(\omega(i))} & \mathbb{P}^{\mathrm{do}(\omega(i))}_Y & \xrightarrow{\mathrm{do}(\omega(j))} & \mathbb{P}^{\mathrm{do}(\omega(j))}_Y
\end{array}
$$

Proof. Let $i, j \in \mathcal{I}_X$ be interventions with $i \leq_X j$. The

---

$\mathbb{P}^{\mathrm{do}(j)}_X$ for some $i \neq_X j$. Doing so would not require a change to Definition 3 or affect the further results of this paper. To avoid notational burden in our exposition, we omit this treatment.

[5]For example, the distribution over the variables X in the observational setting, $\mathbb{P}^\varnothing_X$, changes to $\mathbb{P}^{\mathrm{do}(i)}_X$ if we implement the intervention $\mathrm{do}(i)$, and the partial ordering contains all information about which interventions can be composed.

[6]This problem is elaborated upon in [7].

[7]Since $\omega(i) = \omega(j) \iff \left(\omega(i) \leq_Y \omega(j)\right) \wedge \left(\omega(j) \leq_Y \omega(i)\right)$, which, if the converse held, would imply that $\left(i \leq_X j\right) \wedge \left(j \leq_X i\right)$, which is equivalent to $i = j$.

[8]For instance, if it were necessary that ω be bijective, Theorems 9 and 11 would not hold.

commutativity of the left square of the diagram follows immediately from the definition of an exact transformation. It remains to be shown that the right square of the diagram commutes. By definition we have that $\tau\left(\mathbb{P}_X^{\mathrm{do}(i)}\right) = \mathbb{P}_Y^{\mathrm{do}(\omega(i))}$ and $\tau\left(\mathbb{P}_X^{\mathrm{do}(j)}\right) = \mathbb{P}_Y^{\mathrm{do}(\omega(j))}$. Thus, we only have to show that $\mathbb{P}_Y^{\mathrm{do}(\omega(i))} \leq_Y \mathbb{P}_Y^{\mathrm{do}(\omega(j))}$ as elements of $\mathcal{P}_Y$, i.e. that the arrow $\mathbb{P}_Y^{\mathrm{do}(\omega(i))} \xrightarrow{\mathrm{do}(\omega(j))} \mathbb{P}_Y^{\mathrm{do}(\omega(j))}$ exists. This follows from the order-preservingness of ω. $\square$

4.4 CAUSAL INTERPRETATION OF EXACT TRANSFORMATIONS

The notion of an exact transformation between SEMs was motivated by the desire to analyse the correspondence between two causal models describing the same system at different levels of detail. The purpose of this section is to show that if one SEM can be viewed as an exact transformation of the other, then both can sensibly be thought of as causal models of the same system. In the following, we assume that $\mathcal{M}_Y$ is an exact τ-transformation of $\mathcal{M}_X$ with ω the corresponding map between interventions.

Surjectivity of ω ensures that any intervention in $\mathcal{I}_Y$ can be viewed as an $\mathcal{M}_Y$-level representative of some intervention on the $\mathcal{M}_X$-level. Consequently, if do-interventions on the $\mathcal{M}_X$-level are in correspondence with physical implementations, then surjectivity of ω ensures that do-interventions on the $\mathcal{M}_Y$-level have at least one corresponding physical implementation, i.e. if $\mathcal{M}_X$ is 'rooted in reality', then so is $\mathcal{M}_Y$.

Commutativity of the left hand part of the diagram ensures that the effects of interventions are consistently modelled by $\mathcal{M}_X$ and $\mathcal{M}_Y$. Suppose we want to reason about the effects on the $\mathcal{M}_Y$-level caused by the intervention $j \in \mathcal{I}_Y$. For example, we may wish to reason about how the temperature and pressure of a volume of gaseous particles is affected by being heated. We could perform this reasoning by considering any corresponding $\mathcal{M}_X$-level intervention $i \in \omega^{-1}(\{j\})$ and considering the distribution this implies over $\mathcal{Y}$ via τ. In our example, this would correspond to considering how heating the volume of gas could be modelled by changing the motions of all the gaseous particles and then computing the temperature and pressure of the volume of particles. Commutativity of the left hand part of the diagram implies that $\mathcal{M}_X$ and $\mathcal{M}_Y$ are consistent in the sense that $\mathcal{M}_Y$ allows us to immediately reason about the effect of the intervention $j \in \mathcal{I}_Y$ while being equivalent to performing the steps above. That is, we can reason directly about temperature and pressure when heating a volume of gas without having to perform the intermediate steps that involve the microscopic description of the system.

Commutativity of the right hand side of the diagram en-

sures that once an intervention that fixes a subset of the variables has been performed, we can still consistently reason about the effects of further interventions on the remaining variables in $\mathcal{M}_X$ and $\mathcal{M}_Y$. Furthermore, it ensures that compositionality of do-interventions on the $\mathcal{M}_X$-level carries over to the $\mathcal{M}_Y$-level, i.e. if the intervention j on the $\mathcal{M}_X$-level can be performed additionally to the intervention i in $\mathcal{M}_X$—that is, $i \leq_X j$—, then the same is true of their representations in $\mathcal{M}_Y$.

If $\mathcal{M}_X$ and $\mathcal{M}_Y$ are models of the same system and it has been established that $\mathcal{M}_Y$ is an exact τ-transformation of $\mathcal{M}_X$ for some mapping τ, then the commutativity of the whole diagram in Theorem 6 ensures that they are causally consistent with one another in the sense described in the preceding paragraphs. If we wish to reason about the effects of interventions on the $\mathcal{Y}$-variables then it suffices to use the model $\mathcal{M}_Y$, rather than the (possibly more complex) model $\mathcal{M}_X$. In particular, this means that we can view the $\mathcal{Y}$-variables as causal entities, rather than only functions of underlying 'truly' causal entities. Only if this is the case, causal statements such as 'raising temperature increases pressure' or 'LDL causes heart disease' are meaningful.

4.5 WHAT CAN GO WRONG WHEN A TRANSFORMATION IS NOT EXACT?

In the previous section we argued that our definition of exact transformations between SEMs is a sensible formalisation of causal consistency. In this section we will try to give the reader an intuition for why weakening the conditions of our definition would be problematic. In particular we focus on the requirement that ω be order-preserving, which we view as one of the core ideas of our paper.

The requirement that ω be surjective is, as discussed above, required so that all interventions on the $\mathcal{M}_Y$-level have a corresponding intervention on the $\mathcal{M}_X$-level. If we were to only require that ω be surjective (but not order-preserving), the observational distribution of $\mathcal{M}_X$ may be mapped to an interventional distribution of $\mathcal{M}_Y$, as illustrated by the following example (cf. Figure 2 for an illustration).

Example 7. *Consider the SEM* $\mathcal{M}_X = \{S_X, \mathcal{I}_X, \mathbb{P}_E\}$ *over* $\mathcal{X} = \mathbb{R}^3$ *where*

$$S_X = \left\{X_1 = E_1,\ X_2 = E_2,\ X_3 = X_1 + X_2 + E_3\right\}$$
$$\mathcal{I}_X = \left\{\varnothing,\ \mathrm{do}(X_2 = 0),\ \mathrm{do}(X_1 = 0,\ X_2 = 0)\right\},$$
$$E_1 \sim \mathbb{P}_{E_1},\ E_2 = -E_1,\ E_3 \sim \mathbb{P}_{E_3}$$

where $\mathbb{P}_{E_1}$ *and* $\mathbb{P}_{E_3}$ *are arbitrary distributions. Let* $\tau : \mathcal{X} \to \mathcal{Y} = \mathbb{R}^2$ *be the mapping such that*

$$\tau\left(x_1, x_2, x_3\right) = \left(y_1, y_2\right) = \left(x_1 + x_2, x_3\right)$$

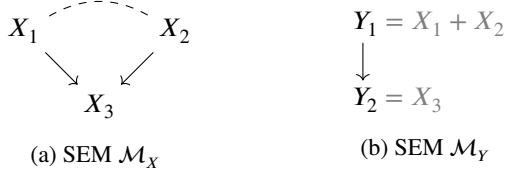(a) SEM $\mathcal{M}_X$ (b) SEM $\mathcal{M}_Y$

Figure 2: Graphical illustration of parent-child relationships for the examples in Section 4.5. The micro-level model $\mathcal{M}_X$ depicted in (a) is to be transformed into the macro-level model $\mathcal{M}_Y$ depicted in (b) which is a coarser descriptions as in it only considers the sum of X_1 and X_2. In Section 4.5 we give examples of what can go wrong if the transformation is not exact.

Let $\mathcal{M}_Y = \{S_Y, \mathcal{I}_Y, \mathbb{P}_F\}$ be an SEM over $\mathcal{Y}$ with

$$S_Y = \left\{Y_1 = F_1,\ Y_2 = Y_1 + F_2\right\}$$
$$\mathcal{I}_Y = \left\{\varnothing,\ \mathrm{do}(Y_1 = 0)\right\},$$
$$F_1 \sim \mathbb{P}_{E_1},\ \ F_2 \sim \mathbb{P}_{E_3}$$

Let $\omega : \mathcal{I}_X \to \mathcal{I}_Y$ be defined by

$$\omega : \begin{cases} \varnothing & \mapsto \mathrm{do}(Y_1 = 0) \\ \mathrm{do}(X_2 = 0) & \mapsto \varnothing \\ \mathrm{do}(X_1 = 0,\ X_2 = 0) & \mapsto \mathrm{do}(Y_1 = 0) \end{cases}$$

Then it is true that $\mathbb{P}^i_{\tau(X)} = \mathbb{P}^{\mathrm{do}(\omega(i))}_Y$ for all $i \in \mathcal{I}_X$, while ω is not order-preserving and $\omega(\varnothing) \neq \varnothing$.

If the SEMs in the above example were used to model the same system, it would be problematic that the observational setting of $\mathcal{M}_X$—a description of the system when not having physically performed any intervention—would correspond to an interventional setting in $\mathcal{M}_Y$, conversely suggesting that the system *had* been intervened upon.

To avoid the above conflict, we could demand in addition to surjectivity that ω map the null intervention of $\mathcal{M}_X$ to the null intervention of $\mathcal{M}_Y$. This additional assumption would ensure commutativity of the left-hand part of the diagram in Theorem 6. However, as the following example shows, this would not ensure that the right-hand part of the diagram commutes for all pairs of interventions $i \leq_X j$, since in this case the arrow from $\mathbb{P}^{\mathrm{do}(\omega(i))}_Y$ to $\mathbb{P}^{\mathrm{do}(\omega(j))}_Y$ may not exist.[9]

Example 8. *Let $\mathcal{X}$, $\mathcal{Y}$ and τ be as in Example 7. Consider the SEM $\mathcal{M}_X = \{S_X, \mathcal{I}_X, \mathbb{P}_E\}$ where*

$$S_X = \left\{X_1 = E_1,\ X_2 = E_2,\ X_3 = X_1 + X_2 + E_3\right\}$$
$$\mathcal{I}_X = \left\{\varnothing,\ \mathrm{do}(X_2 = 0),\ \mathrm{do}(X_1 = 0,\ X_2 = 0)\right\},$$
$$E_1 = 1,\ \ E_2 \sim \mathbb{P}_{E_2},\ \ E_3 \sim \mathbb{P}_{E_3}$$

---

[9]By definition of the poset $\mathcal{P}_Y$, this arrow exists if and only if $\omega(i) \leq_Y \omega(j)$.

where $\mathbb{P}_{E_2}$ and $\mathbb{P}_{E_3}$ are arbitrary distributions. Let $\mathcal{M}_Y = \{S_Y, \mathcal{I}_Y, \mathbb{P}_F\}$ be the SEM over $\mathcal{Y}$ with

$$S_Y = \left\{Y_1 = 1 + F_1,\ Y_2 = Y_1 + F_2\right\}$$
$$\mathcal{I}_Y = \left\{\varnothing,\ \mathrm{do}(Y_1 = 0),\ \mathrm{do}(Y_1 = 1)\right\},$$
$$F_1 \sim \mathbb{P}_{E_2},\ \ F_2 \sim \mathbb{P}_{E_3}$$

Let $\omega : \mathcal{I}_X \to \mathcal{I}_Y$ be defined by

$$\omega : \begin{cases} \varnothing & \mapsto \varnothing \\ \mathrm{do}(X_2 = 0) & \mapsto \mathrm{do}(Y_1 = 1) \\ \mathrm{do}(X_1 = 0,\ X_2 = 0) & \mapsto \mathrm{do}(Y_1 = 0) \end{cases}$$

Then it is true that $\mathbb{P}^i_{\tau(X)} = \mathbb{P}^{\mathrm{do}(\omega(i))}_Y$ for all $i \in \mathcal{I}_X$ and $\omega(\varnothing) = \varnothing$, although ω is not order-preserving.

If the above SEMs were used as models of the same system, they would not suffer from the problem illustrated in Example 7. Suppose now, however, that we have performed the intervention $\mathrm{do}(X_2 = 0)$ in $\mathcal{M}_X$, corresponding to the intervention $\mathrm{do}(Y_1 = 1)$ in $\mathcal{M}_Y$. If we wish to reason about the effect of the intervention $\mathrm{do}(X_1 = 0,\ X_2 = 0)$ in $\mathcal{M}_X$, we run into a problem. $\mathcal{M}_X$ suggests that $\mathrm{do}(X_1 = 0,\ X_2 = 0)$ could be implemented by performing an additional action on top of $\mathrm{do}(X_2 = 0)$. In contrast, $\mathcal{M}_Y$ suggests that implementing the corresponding intervention $\mathrm{do}(Y_1 = 0)$ would conflict with the already performed intervention $\mathrm{do}(Y_1 = 1)$.

5 EXAMPLES OF EXACT TRANSFORMATIONS

In the introduction we motivated the problem considered in this paper by listing three settings in which differing model levels naturally occur. Having now introduced the notion of an exact transformation between SEMs, we provide in this section examples of exact transformations falling into each of these categories. The fact that a single framework can be used to draw an explicit correspondence between differing model levels in each of these settings demonstrates the generality of our framework.

Observe that in each of the following examples, the particular set of interventions considered is important. If we were to allow larger sets of interventions $\mathcal{I}_X$ in the SEM $\mathcal{M}_X$, the transformations given would not be exact. This highlights the importance to the causal modelling process of carefully considering the set of interventions. All proofs are found in the Appendix.

5.1 MARGINALISATION OF VARIABLES

In the following two Theorems we consider two operations that can be performed on SEMs, namely marginalisation of childless or non-intervened variables, and prove

that these are exact transformations. That is, an SEM can be simplified into an SEM with fewer variables by either of these operations without losing any causal content concerning the remaining variables.

Thus if the SEM $\mathcal{M}_Y$ can be obtained from another SEM $\mathcal{M}_X$ by successively performing the operations in the following theorems, then $\mathcal{M}_Y$ is an exact transformation of $\mathcal{M}_X$ and hence the two models are causally consistent. This formally explains why we can sensibly consider causal models that focus on a subsystem $\mathcal{M}_Y$ of a more complex system $\mathcal{M}_X$ (cf. Figure 3). For a measure-theoretic treatment of marginalisation in SEMs, see [3].

Theorem 9 (Marginalisation of childless variables). *Let $\mathcal{M}_X = (S_X, \mathcal{I}_X, \mathbb{P}_E)$ be an SEM and suppose that $\mathbb{I}_Z \subset \mathbb{I}_X$ is a set of indices of variables with no children, i. e. if $i \in \mathbb{I}_Z$ then X_i does not appear in the right-hand side of any structural equation in S_X. Let $\mathcal{Y}$ be the set in which $Y = (X_i : i \in \mathbb{I}_X \setminus \mathbb{I}_Z)$ takes value. Then the transformation $\tau : \mathcal{X} \to \mathcal{Y}$ mapping*

$$\tau : (x_i : i \in \mathbb{I}_X) = x \mapsto y = (x_i : i \in \mathbb{I}_X \setminus \mathbb{I}_Z)$$

naturally gives rise to an SEM $\mathcal{M}_Y$ that is an exact τ-transformation of $\mathcal{M}_X$, corresponding to marginalising out the childless variables X_i for $i \in \mathbb{I}_Z$.

Theorem 10 (Marginalisation of non-intervened variables). *Let $\mathcal{M}_X = (S_X, \mathcal{I}_X, \mathbb{P}_E)$ be an acyclic SEM and suppose that $\mathbb{I}_Z \subset \mathbb{I}_X$ is a set of indices of variables that are not intervened upon by any intervention $i \in \mathcal{I}_X$. Let $\mathcal{Y}$ be the set in which $Y = (X_i : i \in \mathbb{I}_X \setminus \mathbb{I}_Z)$ takes value. Then the transformation $\tau : \mathcal{X} \to \mathcal{Y}$ mapping*

$$\tau : (x_i : i \in \mathbb{I}_X) = x \mapsto y = (x_i : i \in \mathbb{I}_X \setminus \mathbb{I}_Z)$$

naturally gives rise to an SEM $\mathcal{M}_Y$ that is an exact τ-transformation of $\mathcal{M}_X$, corresponding to marginalising out the never-intervened-upon variables X_i for $i \in \mathbb{I}_Z$.

The assumption of acyclicity made in Theorem 10 can be relaxed to allow marginalisation of non-intervened variables in cyclic SEMs, at the expense of extra technical conditions (see Section 3 of [3]).

We remind the reader that our definition of an SEM does not require that the exogenous E-variables be independent. Theorem 10 would not hold if this restriction were made (which is usually the case in the literature); marginalising out a common parent node will in general result in its children having dependent exogenous variables.

5.2 MICRO- TO MACRO-LEVEL

Transformations from micro- to macro-levels may arise in situations in which the micro-level variables can be observed via a 'coarse' measurement device, represented by
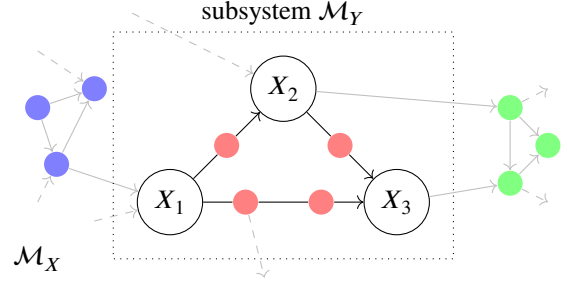

Figure 3: Suppose that there is a complex model $\mathcal{M}_X$ but that we only wish to model the distribution over X_1, X_2, X_3 and how it changes under some interventions on X_1, X_2, X_3. By Theorem 9, we can ignore downstream effects (🟢) after grouping them together as one multivariate variable and by Theorem 10 we can ignore intermediate steps of complex mechanisms (🔴) and treat upstream causes as noise fluctuations (🟣). That is, we can exactly transform the complex SEM $\mathcal{M}_X$ into a simpler model $\mathcal{M}_Y$ by marginalisation.

the function τ, e. g. we can use a thermometer to measure the temperature of a gas, but not the motions of the individual particles. They may also arise due to deliberate modelling choice when we wish to describe a system using higher level features, e. g. viewing the motor cortex as a single entity responsible for movements, rather than as a collection of individual neurons.

In such situations, our framework of exact transformations allows one to investigate whether such a macro-level model admits a causal interpretation. The following theorem provides an exact transformation between a micro-level model $\mathcal{M}_X$ and a macro-level model $\mathcal{M}_Y$ in which the variables are aggregate features of variables in $\mathcal{M}_X$ obtained by averaging (cf. Figure 4).

Theorem 11 (Micro- to macro-level). *Let $\mathcal{M}_X = (S_X, \mathcal{I}_X, \mathbb{P}_{E,F})$ be a linear SEM over the variables $W = (W_i : 1 \leq i \leq n)$ and $Z = (Z_i : 1 \leq i \leq m)$ with*

$$S_X = \{ W_i = E_i : 1 \leq i \leq n \}$$
$$\cup \left\{ Z_i = \sum_{j=1}^{n} A_{ij} W_j + F_i : 1 \leq i \leq m \right\}$$
$$\mathcal{I}_X = \Big\{ \varnothing, \text{do}(W = w), \text{do}(Z = z),$$
$$\text{do}(W = w, Z = z) : w \in \mathbb{R}^n, z \in \mathbb{R}^m \Big\}$$

and $(E, F) \sim \mathbb{P}$ where $\mathbb{P}$ is any distribution over $\mathbb{R}^{n+m}$ and A is a matrix.

Assume that there exists an $a \in \mathbb{R}$ such that each column of A sums to a. Consider the following transformation

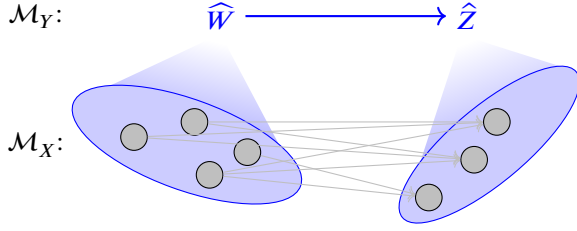$\mathcal{M}_Y$: $\widehat{W} \longrightarrow \widehat{Z}$

$\mathcal{M}_X$:

Figure 4: An illustration of the setting considered in Theorem 11. The micro-variables $W_1, \ldots, W_n$ and $Z_1, \ldots, Z_m$ in the SEM $\mathcal{M}_X$ can be averaged to derive macro-variables $\widehat{W}$ and $\widehat{Z}$ in such a way that the resulting macro-level SEM $\mathcal{M}_Y$ is an exact transformation of the micro-level SEM $\mathcal{M}_X$.

that averages the W and Z variables:

$$\tau : \mathcal{X} \to \mathcal{Y} = \mathbb{R}^2$$

$$\begin{pmatrix} W \\ Z \end{pmatrix} \mapsto \begin{pmatrix} \widehat{W} \\ \widehat{Z} \end{pmatrix} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} W_i \\ \frac{1}{m} \sum_{j=1}^{m} Z_j \end{pmatrix}$$

Further, let $\mathcal{M}_Y = \left(S_Y, \mathcal{I}_Y, \mathbb{P}_{\widehat{E}, \widehat{F}} \right)$ over the variables $\left\{ \widehat{W}, \widehat{Z} \right\}$ be an SEM with

$$S_Y = \left\{ \widehat{W} = \widehat{E}, \ \widehat{Z} = \frac{a}{m} \widehat{W} + \widehat{F} \right\}$$

$$\mathcal{I}_Y = \left\{ \varnothing, \ \mathrm{do}(\widehat{W} = \widehat{w}), \ \mathrm{do}(\widehat{Z} = \widehat{z}), \right.$$

$$\left. \mathrm{do}(\widehat{W} = \widehat{w}, \widehat{Z} = \widehat{z}) : \widehat{w} \in \mathbb{R}, \ \widehat{z} \in \mathbb{R} \right\}$$

$$\widehat{E} \sim \frac{1}{n} \sum_{i=1}^{n} E_i, \quad \widehat{F} \sim \frac{1}{m} \sum_{i=1}^{m} F_i$$

Then $\mathcal{M}_Y$ is an exact τ-transformation of $\mathcal{M}_X$.

5.3 STATIONARY BEHAVIOUR OF DYNAMICAL PROCESSES

In this section we provide an example of an exact transformation between an SEM $\mathcal{M}_X$ describing a time-evolving system and another SEM $\mathcal{M}_Y$ describing the system after it has equilibrated. In this setting, τ could be thought of as representing our ability to only measure the time-evolving system at a single point in time, after the transient dynamics have taken place.

In particular, we consider a discrete-time linear dynamical system with identical noise and provide the explicit form of an SEM that models the distribution of the equilibria under each intervention (cf. Figure 5).[10]

---

[10]Note that the assumption that the transition dynamics be

Theorem 12 (Discrete-time linear dynamical process with identical noise). *Let* $\mathcal{M}_X = \left(S_X, \mathcal{I}_X, \mathbb{P}_E \right)$ *over the variables* $\left\{ X_t^i \ : \ t \in \mathbb{Z}, \ i \in \{1, \ldots, n\} \right\}$ *be a linear SEM with*

$$S_X = \left\{ X_{t+1}^i = \sum_{j=1}^{n} A_{ij} X_t^j + E_t^i \ : \ i \in \{1, \ldots, n\}, t \in \mathbb{Z} \right\}$$

$$i.\, e.\ X_{t+1} = A X_t + E_t$$

$$\mathcal{I}_X = \left\{ \mathrm{do}(X_t^j = x_j \ \forall t \in \mathbb{Z}, \forall j \in J) : x \in \mathbb{R}^{|J|}, \ J \subseteq \{1, \ldots, n\} \right\}$$

$$E_t = E \ \forall t \in \mathbb{Z} \ \text{where} \ E \sim \mathbb{P}$$

where $\mathbb{P}$ *is any distribution over* $\mathbb{R}^n$ *and* A *is a matrix.*

Assume that the linear mapping $v \mapsto Av$ *is a contraction. Then the following transformation is well-defined under any intervention* $i \in \mathcal{I}_X$:[11]

$$\tau : \mathcal{X} \to \mathcal{Y}$$

$$(x_t)_{t \in \mathbb{Z}} \mapsto y = \lim_{t \to \infty} x_t$$

Let $\mathcal{M}_Y = \left(S_Y, \mathcal{I}_Y, \mathbb{P}_F \right)$ *be the (potentially cyclic) SEM over the variables* $\left\{ Y^i \ : \ i \in \{1, \ldots, n\} \right\}$ *with*

$$S_Y = \left\{ Y^i = \frac{\sum_{j \neq i} A_{ij} Y^j}{1 - A_{ii}} + \frac{F^i}{1 - A_{ii}} \ : \ i \in \{1, \ldots, n\} \right\}$$

$$\mathcal{I}_Y = \left\{ \mathrm{do}(Y^j = y_j \ \forall j \in J) : y \in \mathbb{R}^{|J|}, \ J \subseteq \{1, \ldots, n\} \right\}$$

$$F \sim \mathbb{P}$$

Then $\mathcal{M}_Y$ *is an exact* τ-*transformation of* $\mathcal{M}_X$.

The above theorem demonstrates how a linear additive SEM can arise as a result of making observations of a dynamical process. This supports one interpretation of SEMs as a description of a dynamical process that equilibrates quickly compared to its external environment.[12] The framework of exact transformations allows us to explain in a precise way the sense in which such equilibrium models can be used as causal descriptions of an underlying dynamical process.

---

linear can be relaxed to more general non-linear mappings. In this case, however, the structural equations of $\mathcal{M}_Y$ can only be written in terms of implicit solutions to the structural equations of $\mathcal{M}_X$. For purposes of exposition, we stick here to the simpler case of linear dynamics.

[11]In Appendix D.1 we show that A being a contraction mapping ensures that the sequence $(X_t)_{t \in \mathbb{Z}}$ defined by $\mathcal{M}_X$ converges everywhere under any intervention $i \in \mathcal{I}_X$. That is, for any realisation $(x_t)_{t \in \mathbb{Z}}$ of this sequence, its limit $\lim_{t \to \infty} x_t$ as a sequence of elements of $\mathbb{R}^n$ exists.

[12]This interpretation corresponds to the assumption that the noise in the dynamical model is constant through time, and is used by e.g. [10, 12, 14, 15] and [13] to meaningfully interpret cyclic SEMs.
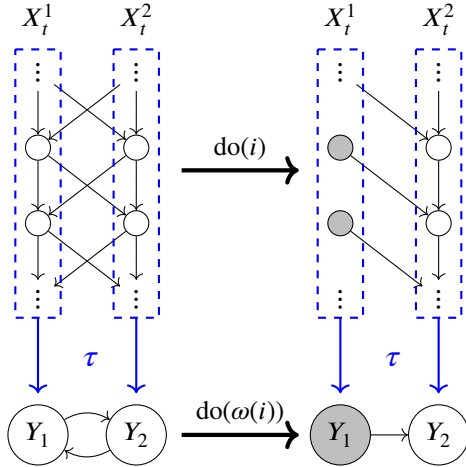
Figure 5: An illustration of the setting considered in Theorem 12. The discrete-time dynamical process is exactly transformed into a model describing its equilibria.

This result also sheds light on the interpretation of cyclic causal models. One interpretation of the structural equations of an acyclic SEM is that they represent a temporally ordered series of mechanisms by which data are generated. This is not possible in the case that the SEM exhibits cycles: there does not exist a partial ordering on the variables and hence one cannot think of each variable being generated temporally downstream of its parents. By showing that cyclic SEMs can arise as exact transformations of *acyclic* SEMs, we provide an interpretation of cyclic SEMs that does not suffer from the above problem.

6 DISCUSSION AND FUTURE WORK

It's turtles all the way down! There is no such thing as a 'correct' model, but in this paper we introduced the notions of exact transformations between SEMs to evaluate when two SEMs can be viewed as causally consistent models of the same system. Illustrating how these notions can be used in order to relate differing model levels, we proved in Section 5 the exactness of transformations occurring in three different settings. These have implications for the following questions in causal modelling: When can we model only a subsystem of a more complex system? When does a micro-level system admit a causal description in terms of macro-level features? How do cyclic causal models arise?

Our work has implications for other problems in causal modelling. It suggests that ambiguous manipulations [18] may be thought of as arising due to the application of an in-exact transformation to an SEM $\mathcal{M}_X$. This was illustrated in Section 1.1 in which LDL and HDL cholesterol were only measured via their sum TC, resulting in a model that

suffered from the problem of ambiguous manipulations (cf. Figure 1b) since it was not an exact transformation of the underlying model (cf. Figure 1a). This is related to the problem of causal variable definition as studied by [7].

A future line of enquiry would be to generalise the notion of an exact transformation in order to analyse the trade-off between model accuracy and model complexity for causal modelling using SEMs. For a transformation to be exact, we require that the posets $\mathcal{P}_{\tau(X)}$ and $\mathcal{P}_Y$ be equal. One could imagine a 'softening' of this requirement such that the distributions in the posets are required to be only approximately equal. A slightly inaccurate model with a small number of variables may be preferable to an accurate but complex model.

We discussed the importance of an order-preserving ω to ensure a notion of causal consistency between two SEMs. It would be interesting to better understand the conditions under which different properties of consistency between causal models hold – for instance, counterfactual reasoning, which we have not discussed in this paper.

While we have introduced the notion of an exact transformation, we have not provided any criterion to choose from amongst the set of all possible exact transformations of an SEM. Foundational work in a similar direction to ours has been done by [4, 5], who consider a particular discrete setting. They provide algorithms to learn a transformation of a micro-level model to a macro-level model with desirable information-theoretic properties. We conjecture that our framework may lead to extensions of their work, e. g. to the continuous setting.

Finally, suppose that we have made observations of an underlying system $\mathcal{M}_X$ via a measurement device τ, and that we want to fit an SEM $\mathcal{M}_Y$ from a restricted model class to our data. By using our framework, asking whether or not $\mathcal{M}_Y$ admits a causal interpretation consistent with $\mathcal{M}_X$ reduces to asking whether the transformation is exact. More generally, by fixing any two of $\mathcal{M}_X$, τ and $\mathcal{M}_Y$, we can ask what properties must be fulfilled by the third in order for the two models to be causally consistent. We hope that this may lead to the practical use of SEMs being theoretically grounded.

Acknowledgements

References

[1] R. Balian. *From microphysics to macrophysics*. Springer, 1992.

[2] K. A. Bollen. *Structural equations with latent variables*. John Wiley & Sons, 2014.

[3] S. Bongers, J. Peters, B. Schölkopf, and J. M. Mooij. Structural causal models: Cycles, marginalizations, exogenous reparametrizations and reductions. *arXiv preprint arXiv:1611.06221*, 2016.

[4] K. Chalupka, P. Perona, and F. Eberhardt. Visual causal feature learning. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pages 181–190. AUAI Press, 2015.

[5] K. Chalupka, P. Perona, and F. Eberhardt. Multi-level cause-effect systems. In *The 19th International Conference on Artificial Intelligence and Statistics.*, 2016.

[6] D. Dash and M. J. Druzdzel. Caveats for causal reasoning with equilibrium models. *Lecture notes in computer science*, pages 192–203, 2001.

[7] F. Eberhardt. Green and grue causal variables. *Synthese*, 193(4):1029–1046, 2016.

[8] F. M. Fisher. A correspondence principle for simultaneous equation models. *Econometrica: Journal of the Econometric Society*, pages 73–92, 1970.

[9] E. P. Hoel, L. Albantakis, and G. Tononi. Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences*, 110(49):19790–19795, 2013.

[10] A. Hyttinen, F. Eberhardt, and P. O. Hoyer. Learning linear cyclic causal models with latent variables. *Journal of Machine Learning Research*, 13(Nov): 3387–3439, 2012.

[11] Y. Iwasaki and H. A. Simon. Causality and model abstraction. *Artificial Intelligence*, 67(1):143–194, 1994.

[12] G. Lacerda, P. L. Spirtes, J. Ramsey, and P. O. Hoyer. Discovering cyclic causal models by independent components analysis. In *Proceedings of the Twenty-Fourth Conference Annual Conference on Uncertainty in Artificial Intelligence*, 2008.

[13] J. M. Mooij and T. Heskes. Cyclic causal discovery from continuous equilibrium data. In *Proceedings of the Twenty-Ninth Conference Annual Conference on Uncertainty in Artificial Intelligence*, 2013.

[14] J. M. Mooij, D. Janzing, T. Heskes, and B. Schölkopf. On causal discovery with cyclic additive noise models. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 24 (NIPS*2011)*, pages 639–647, 2011.

[15] J. M. Mooij, D. Janzing, and B. Schölkopf. From Ordinary Differential Equations to Structural Causal Models: the deterministic case. In *Proceedings of the Twenty-Ninth Conference Annual Conference on Uncertainty in Artificial Intelligence*, pages 440–448, 2013.

[16] J. Pearl. *Causality*. Cambridge University Press, 2009.

[17] H. A. Simon and A. Ando. Aggregation of variables in dynamic systems. *Econometrica: journal of the Econometric Society*, pages 111–138, 1961.

[18] P. Spirtes and R. Scheines. Causal inference of ambiguous manipulations. *Philosophy of Science*, 71 (5):833–845, 2004.

[19] P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, prediction, and search*. MIT press, 2000.

[20] D. Steinberg. *The Cholesterol Wars: The Skeptics vs the Preponderance of Evidence*. Academic Press, 2011.

[21] A. S. Truswell. *Cholesterol and beyond: the research on diet and coronary heart disease 1900-2000*. Springer Science & Business Media, 2010.

Appendix

A   PROOFS FOR SECTION 4.3: ELEMENTARY EXACT TRANSFORMATIONS

Lemma 4. *The identity mapping and permuting the labels of variables are both exact transformations. That is, if $\mathcal{M}_X$ is an SEM and $\pi : \mathbb{I}_X \to \mathbb{I}_X$ is a bijection then the transformation*

$$\tau : \mathcal{X} \to \mathcal{Y}$$
$$(x_i : i \in \mathbb{I}_X) \mapsto (x_{\pi(i)} : i \in \mathbb{I}_X)$$

naturally gives rise to an SEM $\mathcal{M}_Y$ that is an exact τ-transformation of $\mathcal{M}_X$, corresponding to relabelling the variables.

Proof of Lemma 4. Consider the SEM $\mathcal{M}_Y$ obtained from $\mathcal{M}_X$ by replacing, for all $i \in \mathbb{I}_X$, any occurrence of X_i in the structural equations S_X and interventions $\mathcal{I}_X$ by $Y_{\pi(i)}$ and leaving the distribution over the exogenous variables unchanged. □

Proof of Lemma 5 (Transitivity of exact transformations). Let $\omega_{ZY} : \mathcal{I}_Y \to \mathcal{I}_Z$ and $\omega_{YX} : \mathcal{I}_X \to \mathcal{I}_Y$ be the mappings between interventions corresponding to the exact transformations τ_{ZY} and τ_{YX} respectively and define $\omega_{ZX} = \omega_{ZY} \circ \omega_{YX} : \mathcal{I}_X \to \mathcal{I}_Z$. Then ω_{ZX} is surjective and order-preserving since both ω_{ZY} and ω_{YX} are surjective and order-preserving. Since τ_{ZY} and τ_{YX} are exact it follows that for all $i \in \mathcal{I}_X$

$$\mathbb{P}^i_{\tau_{ZX}(X)} = \mathbb{P}^{\omega_{ZY}(\omega_{YX}(i))}_{\tau_{ZY}(\tau_{YX}(X))} = \mathbb{P}^{\mathrm{do}(\omega_{ZX}(i))}_Z$$

i. e. $\mathcal{M}_Z$ is an τ_{ZX}-exact transformation of $\mathcal{M}_X$. □

B   PROOFS FOR SECTION 5.1: MARGINALISATION OF VARIABLES

Proof of Theorem 9 (Marginalisation of childless variables). By Lemma 5 it suffices to proof this for marginalisation of one childless variable. Without loss of generality, let X_1 be the childless variable to be marginalised out.

Let $\mathcal{M}_Y = (S_Y, \mathcal{I}_Y, \mathbb{P}_F)$ be the SEM where

- the structural equations S_Y are obtained from S_X by removing the structural equation corresponding to the childless variable X_1;

- $\mathcal{I}_Y$ is the image of the map $\omega : \mathcal{I}_X \to \mathcal{I}_Y$ that drops any reference to the variable X_1 (e. g. $\mathrm{do}(X_1 = x_1, X_2 = x_2) \in \mathcal{I}_X$ would be mapped to $\mathrm{do}(X_2 = x_2) \in \mathcal{I}_Y$);

- $F = (E_i : i \in \mathbb{I}_X \setminus \{1\})$ are the remaining noise variables distributed according to their marginal distribution under $\mathbb{P}_E$.

By construction, ω is surjective and order-preserving. Let $i \in \mathcal{I}_X$ be any intervention. The variable X_1 being childless ensures that the law on the remaining variables $X_k, k \in \mathbb{I}_X \setminus \{1\}$ that we obtain by *marginalisation* of the childless variable, i. e. $\mathbb{P}^i_{\tau(X)}$, is equivalent to the law one obtains by simply *dropping* the childless variable, which is exactly what the law under $\mathcal{M}_Y$ amounts to, i. e. $\mathbb{P}^{\omega(\mathrm{do}(i))}_Y$. □

Proof of Theorem 10 (Marginalisation of non-intervened variables). By Lemma 5 it suffices to proof this for marginalisation of one never-intervened-upon variable. Without loss of generality, let X_1 be the never-intervened-upon variable to be marginalised out. By acyclicity of the SEM $\mathcal{M}_X$, the structural equation corresponding to variable X_1 is of the form $X_1 = f_1\left(\mathbf{X}_{\mathrm{pa}(1)}, E_1\right)$ and X_1 does not appear in the structural equation for any of its ancestors.

Now let $\mathcal{M}_Y = (S_Y, \mathcal{I}_Y, \mathcal{P}_F)$ be the SEM where

- $\mathcal{I}_Y = \mathcal{I}_X$;

- $F_i = ((E_i, E_1) \, : \, i \in \mathbb{I}_X \setminus \{1\})$ are the noise variables distributed as implied by $\mathbb{P}_E$;

- the structural equations $\mathcal{S}_Y$ are obtained from $\mathcal{S}_X$ by removing the structural equation of X_1 and replacing any occurrence of X_1 in the right-hand side of the structural equations of children of X_1 by $f_1\left(\mathbf{X}_{\mathrm{pa}(1)}, E_1\right)$, yielding
$$X_i = f_i\left(f_1\left(\mathbf{X}_{\mathrm{pa}(1)}, E_1\right), \, \mathbf{X}_{\mathrm{pa}(i)}, \, E_i\right).$$

Note that the structural equations of the resulting SEM are still acyclic and are all of the form $X_i = h_i\left(\mathbf{X}_{\setminus i}, \, F_i\right)$.

Then $\mathcal{M}_Y$ is, by construction, an τ-exact transformation of $\mathcal{M}_X$ for $\omega = \mathrm{id}$. $\qquad \square$

C   PROOF FOR SECTION 5.2: MICRO- TO MACRO-LEVEL

Proof of Theorem 11. We begin by defining a mapping between interventions

$$\omega : \mathcal{I}_X \to \mathcal{I}_Y$$
$$\varnothing \mapsto \varnothing$$
$$\mathrm{do}(W = w) \mapsto \mathrm{do}\left(\widehat{W} = \frac{1}{n}\sum_{i=1}^{n} w_i\right)$$
$$\mathrm{do}(Z = z) \mapsto \mathrm{do}\left(\widehat{Z} = \frac{1}{m}\sum_{i=1}^{m} z_i\right)$$
$$\mathrm{do}(W = w, Z = z) \mapsto \mathrm{do}\left(\widehat{W} = \frac{1}{n}\sum_{i=1}^{n} w_i, \, \widehat{Z} = \frac{1}{m}\sum_{i=1}^{m} z_i\right)$$

Note that ω is surjective and order-preserving (in fact, it is an order embedding). Therefore, it only remains to show that the distributions implied by $\tau(X)$ under any intervention $i \in \mathcal{I}_X$ agree with the corresponding distributions implied by $\mathcal{M}_Y$. That is, we have to show that
$$\mathbb{P}^i_{\tau(X)} = \mathbb{P}^{\mathrm{do}(\omega(i))}_Y \quad \forall i \in \mathcal{I}_X$$
In the observational setting, the distribution over $\mathcal{Y}$ is implied by the following equations:

$$\widehat{W} = \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{n}\sum_{i=1}^{n} E_i$$
$$\widehat{Z} = \frac{1}{m}\sum_{i=1}^{m} Z_i = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij} W_j + F_i\right) = \frac{a}{m}\widehat{W} + \frac{1}{m}\sum_{i=1}^{m} F_i$$

Since the distributions of the exogenous variables in $\mathcal{M}_Y$ are given by $\widehat{E} \sim \frac{1}{n}\sum_{i=1}^{n} E_i$, $\widehat{F} \sim \frac{1}{m}\sum_{i=1}^{m} F_i$, it follows that $\mathbb{P}^{\mathrm{do}(\varnothing)}_{\tau(X)}$ and $\mathbb{P}^{\mathrm{do}(\varnothing)}_Y$ agree. Similarly, the push-forward measure on $\mathcal{Y}$ induced by the intervention $\mathrm{do}(W = w) \in \mathcal{I}_X$ is given by

$$\widehat{W} = \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{n}\sum_{i=1}^{n} w_i$$
$$\widehat{Z} = \frac{1}{m}\sum_{i=1}^{m} Z_i = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij} W_j + F_i\right) = \frac{a}{m}\widehat{W} + \frac{1}{m}\sum_{i=1}^{m} F_i$$

which is the same as the distribution induced by the ω-corresponding intervention $\mathrm{do}\left(\widehat{W} = \frac{1}{n}\sum_{i=1}^{n} w_i\right)$ in $\mathcal{M}_Y$.

Similar reasoning shows that this also holds for the interventions $\mathrm{do}(Z = z)$ and $\mathrm{do}(W = w, Z = z)$.

$\qquad \square$

D   PROOF FOR SECTION 5.3: STATIONARY BEHAVIOUR OF DYNAMICAL PROCESSES

Proof of Theorem 12. We begin by defining a mapping between interventions

$$\omega : \mathcal{I}_X \to \mathcal{I}_Y$$
$$\mathrm{do}(X_t^j = x_j \ \forall t \in \mathbb{Z}, \ \forall j \in J) \mapsto \mathrm{do}(Y^j = x_j \ \forall j \in J)$$

Note that ω is surjective and order-preserving (in fact, it is an order embedding). Therefore, it only remains to show that the distributions implied by $\tau(X)$ under any intervention $i \in \mathcal{I}_X$ agree with the corresponding distributions implied by $\mathcal{M}_Y$. That is, we have to show that

$$\mathbb{P}^i_{\tau(X)} = \mathbb{P}_Y^{\mathrm{do}(\omega(i))} \quad \forall i \in \mathcal{I}_X$$

For this we consider, without loss of generality, the distribution arising from performing the $\mathcal{M}_X$-level intervention

$$i = \mathrm{do}(X_t^j = x_j \ \forall t \in \mathbb{Z}, \forall j \le m \le n) \in \mathcal{I}_X$$

for $m \in [n]$ (for $m = 0$ this amounts to the null-intervention).

Since A is a contraction mapping, it follows from Lemma 15 that for any intervention in $\mathcal{I}_X$, the sequence of random variables X_t defined by $\mathcal{M}_X$ converges everywhere. That is, there exists a random variable X_* such that $X_t \xrightarrow[t \to \infty]{\text{everywhere}} X_*$. In the case of the intervention i above, the random variable X_* satisfies:

$$\begin{cases} X_*^k = x_k & \text{if } k \le m \\ X_*^k = \sum_j A_{kj} X_*^j + E^k & \text{if } m < k \le n \end{cases} \tag{1}$$

Since $\tau(X) = \lim_{t \to \infty} X_t$, it follows from the definition of X_* that $\tau(X) = X_*$, and hence $\tau(X)$ also satisfies the equations above. It follows (rewriting the second line in Equation 1 above) that under the push-forward measure $\mathbb{P}^i_{\tau(X)} = \tau\left(\mathbb{P}_X^{\mathrm{do}(i)}\right)$ the distribution of the random variable $\tau(X) = X_*$ is given by:

$$\begin{cases} X_*^k = x_k & \text{if } k \le m \\ X_*^k = \dfrac{\sum_{j \ne k} A_{kj} X_*^j}{1 - A_{kk}} + \dfrac{E^k}{1 - A_{kk}} & \text{if } m < k \le n \end{cases}$$

We need to compare this to the law of Y as implied by $\mathcal{M}_Y$ under the intervention $\omega(i)$, i. e. $\mathbb{P}_Y^{\mathrm{do}(\omega(i))}$. The $\mathcal{M}_Y$-level intervention $\omega(i)$ corresponding to i is

$$\omega(i) = \mathrm{do}(Y^j = x_j \ \forall j \le m \le n) \in \mathcal{I}_Y$$

and so the structural equations of $\mathcal{M}_Y$ under the intervention $\omega(\mathrm{do}(i))$ are

$$\begin{cases} Y^k = x_k & \text{if } k \le m \\ Y^k = \dfrac{\sum_{j \ne k} A_{kj} Y^j}{1 - A_{kk}} + \dfrac{F^k}{1 - A_{kk}} & \text{if } m < k \le n \end{cases}$$

Since $F \sim E$ it indeed follows that $\tau(X) \sim Y$, i. e. $\mathbb{P}^i_{\tau(X)} = \mathbb{P}_Y^{\mathrm{do}(\omega(i))}$.

Thus $\mathcal{M}_Y$ is an exact τ-transformation of $\mathcal{M}_X$. $\square$

D.1   CONTRACTION MAPPING AND CONVERGENCE

The following Lemmata show that A being a contraction mapping ensures that the sequence $(X_t)_{t \in \mathbb{Z}}$ defined by $\mathcal{M}_X$ in Theorem 12 converges everywhere under any intervention $i \in \mathcal{I}_X$. That is, for any realisation $(x_t)_{t \in \mathbb{Z}}$ of this sequence, its limit $\lim_{t \to \infty} x_t$ as a sequence of elements of $\mathbb{R}^n$ exists.

Lemma 13. *Suppose that the function*

$$f : \mathbb{R}^n \to \mathbb{R}^m$$
$$x \mapsto f(x)$$

is a contraction mapping. Then, for any $e \in \mathbb{R}^m$, so is the function

$$f^* : \mathbb{R}^n \to \mathbb{R}^m$$
$$x \mapsto f(x) + e$$

Proof. By definition, there exists $c < 1$ such that for any $x, y \in \mathbb{R}^n$,

$$\|f^*(x) - f^*(y)\| = \|(f(x) + e) - (f(y) + e)\| = \|f(x) - f(y)\| \leq c\|x - y\|$$

and hence f^* is a contraction mapping. $\qquad\square$

Lemma 14. *Suppose that the function*

$$f : \mathbb{R}^n \to \mathbb{R}^n$$
$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}$$

is a contraction mapping. Then for any $m \leq n$, and $x_i^* \in \mathbb{R}$, $i \in [m]$, so is the function

$$f^* : \mathbb{R}^n \to \mathbb{R}^n$$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} x_1^* \\ \vdots \\ x_m^* \\ f_{m+1}(x) \\ \vdots \\ f_n(x) \end{pmatrix}$$

Proof. By definition, there exists $c < 1$ such that for any $x, y \in \mathbb{R}^n$,

$$\|f^*(x) - f^*(y)\| = \left\| \begin{pmatrix} x_1^* \\ \vdots \\ x_m^* \\ f_{m+1}(x) \\ \vdots \\ f_n(x) \end{pmatrix} - \begin{pmatrix} x_1^* \\ \vdots \\ x_m^* \\ f_{m+1}(y) \\ \vdots \\ f_n(y) \end{pmatrix} \right\| = \left\| \begin{pmatrix} 0 \\ \vdots \\ 0 \\ f_{m+1}(x) - f_{m+1}(y) \\ \vdots \\ f_n(x) - f_n(y) \end{pmatrix} \right\| \leq \left\| \begin{pmatrix} f_1(x) - f_1(y) \\ \vdots \\ f_n(x) - f_n(y) \end{pmatrix} \right\|$$

$$= \|f(x) - f(y)\|$$
$$\leq c\|x - y\|$$

and hence f^* is a contraction mapping. $\qquad\square$

Lemma 15. *Consider the SEM $\mathcal{M}_X$ in Theorem 12, and suppose that the linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ is a contraction mapping. Then, for any intervention $i \in \mathcal{I}_X$, the sequence of X_t converges everywhere.*

Proof. Consider, without loss of generality, the intervention

$$\text{do}(X_t^j = x_j \ \forall t \in \mathbb{Z}, \forall j \leq m \leq n) \in \mathcal{I}_X$$

for $m \in [n]$ (for $m = 0$ this amounts to the null-intervention). The structural equations under this intervention are

$$\begin{cases} X_{t+1}^k = x_k & \text{if } k \leq m \\ X_{t+1}^k = \sum_j A_{kj} X_t^j + E^k & \text{if } m < k \leq n \end{cases}$$

and thus the sequence X_t can be seen to transition according to the function $f = g \circ h$, where

$$h : \mathbb{R}^n \to \mathbb{R}^n$$
$$v \mapsto w = Av + E$$

$$g : \mathbb{R}^n \to \mathbb{R}^n$$

$$w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ w_{m+1} \\ \vdots \\ w_n \end{pmatrix}$$

By Lemma 13 and Lemma 14, f is a contraction mapping for any fixed E. Thus, by the contraction mapping theorem, the sequence of X_t converges everywhere to a unique fixed point. $\square$