Meaningful Causal Aggregation and Paradoxical Confounding

Yuchen Zhu   YUCHEN.ZHU.18@UCL.AC.UK
Department of Computer Science, University College London, UK

Kailash Budhathoki   KAIBUD@AMAZON.DE

Jonas Kübler   KUEBJ@AMAZON.DE

Dominik Janzing   JANZIND@AMAZON.DE
Amazon Research Tübingen, Germany

Editors: Francesco Locatello and Vanessa Didelez

Abstract

In aggregated variables the impact of interventions is typically ill-defined because different micro-realizations of the same macro-intervention can result in different changes of downstream macro-variables. We show that this ill-definedness of causality on aggregated variables can turn unconfounded causal relations into confounded ones and vice versa, depending on the respective micro-realization. We argue that it is practically infeasible to only use aggregated causal systems when we are free from this ill-definedness. Instead, we need to accept that macro causal relations are typically defined only with reference to the micro states. On the positive side, we show that cause-effect relations can be aggregated when the macro interventions are such that the distribution of micro states is the same as in the observational distribution; we term this natural macro interventions. We also discuss generalizations of this observation.

Keywords: Causality, Causal Abstraction

1. Introduction

High-stake topics in economics, politics, and health frequently hinge on aggregated variables like *employment rate*, *voter participation*, and *vaccination rate*. The aggregations often encompass entire business sectors or populations of a region. This prevalence motivates comprehending the aggregation of causal relations between such variables. Drawing from Chalupka et al. (2016); Rubenstein et al. (2017), we term these aggregates as *macro-variables* and the fine-grained variables as *micro-variables*. We first identify three challenges of aggregating causal relations.

Challenge 1. Aggregated causal relations can be ambiguous. As shown in prior studies (Rubenstein et al., 2017; Spirtes and Scheines, 2004; Beckers and Halpern, 2019), consider the medical case of total cholesterol's (TC) influence on heart disease (HD). Here, TC, comprised of low-density lipoprotein (LDL) and high-density lipoprotein (HDL), is a macro-variable. Conflicting findings arise on TC's effect on HD risks. The discrepancy stems from LDL and HDL having inverse impacts on HD, leading to varied interpretations of TC based on LDL and HDL compositions.

Rubenstein et al. (2017) introduce exact transformation between macro- and micro-models. Yet, their criteria are stringent, demanding that every *allowed* microscopic realization of a macro-intervention equally affects the aggregated downstream variables, which is a strong assumption on the set of interventions allowed. Beckers and Halpern (2019) extend this to even more robust consistency definitions, where equal effects are even required of *all* potential microscopic realizations of the same macro-intervention.

In cases impossible for exact transformation, Beckers et al. (2019) introduce a family of methods to measure the difference between macro- and micro-models. However, this requires choosing weights over interventions on micro-variables. Given the high-dimensionality of micro-variables and the expansive realm of potential interventions, reaching a consensus on the weights is challenging for practitioners. Moreover, their discussion cannot be easily extended to address the subsequent challenge.

Challenge 2. Bridging macro intervention to micro implementation.   Practitioners need to realize high-level interventions on the low level, like actualizing a GDP boost and seeing how that impacts civil conflict (Miguel et al., 2004). One might adapt the idea of pre-specifying a distribution over micro-interventions from Beckers et al. (2019, Definition 4.7), where a macro-intervention would be associated with a distribution over a family of micro-interventions. However, two issues arise: firstly, gaining consensus on the distribution, as highlighted in Challenge 1, is tough. Secondly, this yields a spread of micro-interventions that may not individually align with the intended macro-outcome, even if they do on average. The authors offer no clear guidance on how to actualize micro-interventions while maintaining consistency with the macro-level.

Challenge 3. The micro-model may not be known.   In some scenarios, only isolated facts about the micro-model are known. For example, one may know that the relation between product sales and the revenue generated are not confounded, but the exact micro-level causal model is not known. In this case, how should we implement a high-level intervention on the low level such that the high-level causal effects implied by the high-level causal model are realized?

Driven by the three challenges, we delve into causal claims of interventions on macro-variables, *relative to their implementations on the micro level*. To the best of our knowledge, this is a novel perspective of aggregating micro causal models. Janzing and Garrido Mejia (2022) also explores ill-definedness of causality arising from ill-definedness of interventions, but not in the specific sense of aggregating micro-models. We make the following contributions: 1) Beyond the existing discussion on ambiguity of macro-interventions, we uncover that even the concept of *confoundedness* is ill-defined without an appropriate treatment of causal aggregation. Given the central role of confoundedness in causal inference, we hope this will intensify existing interest in causal aggregation. 2) On small micro-level causal graphs, we propose the concepts of 'confounding-inhibiting micro-realizations', which induces unconfoundedness in the macro-model with an appropriately defined micro-level implementation, and 'natural micro-realization', which allows the practitioner to preserve unconfoundedness of the micro-level but without knowledge of the micro-causal model. 3) We extend the concept of 'natural intervention' to larger graphs, and propose macro-backdoor adjustment.

2. Preliminaries: Macro- and micro-intervention

In this work, micro-level causal models are standard acyclic structural equation models, introduced by Pearl (2009) and using notation from Peters et al. (2017):

Definition 1   *A **micro causal model** is an acyclic structural causal model (SCM). A **structural causal model (SCM)** $M := (\mathbf{S}, P_\mathbf{N})$ consists of a collection $\mathbf{S}$ of d **(structural)** assignments*

$$X_j := f_j(\mathbf{PA}_j, U_j), \ \ j = 1, ..., d, \tag{1}$$

*where $\mathbf{PA}_j \subseteq \{X_1, \cdots, X_d\} \setminus \{X_j\}$ are called **parents of** X_j; and a joint distribution $P_\mathbf{U} = P_{U_1, \cdots, U_d}$ over the noise variables, which we require to be jointly independent. The corresponding*

causal DAG G has nodes X_j and arrows into X_j from each of its parents. Note, that X_j does not have to be one-dimensional.

Macro-variables arise from applying an aggregation map on the micro-variables:

Definition 2 (aggregation map, macro-variables) *Given a random variable X with range $\mathcal{X}$, an aggregation map is a surjective but non-injective function $\pi : \mathcal{X} \to \bar{\mathcal{X}}$. Define $\bar{X} = \pi(X)$ to be the macro-variable of X under π.*

Graphically, we represent aggregation by the graph $X \leftrightarrow \bar{X}$, where the bi-directed arrow visualizes the fact that $\bar{X}$ can be seen as effect of X in the observational distribution, but the direction is reversed once we talk about interventions on $\bar{X}$. Intuitively, readers can think about total sales as a consequence of individual sales before intervention; during an intervention on total sales, its change is reflected in individual-level sales (so the arrow is reversed)[1].

We will later consider separate aggregations at each potentially high-dimensional node of the micro-causal model. Each X_j will often be a vector in $\mathbb{R}^n$ and $\bar{X}_j$ the sum or average of all its components. In our examples we will mostly work with the sum aggregation map, so we use the same symbol π instead of π_j for each node. Given a distribution $P(X_1, \ldots, X_n)$ over the micro-variables, defining π_j for each node induces a joint distribution $P(X_1, \ldots, X_n, \pi(X_1,), \ldots, \pi(X_n))$.

Definition 3 (amalgamated graph) *Given a micro causal model and a set of aggregation maps, we define the amalgamated graph, $G^a := (\mathcal{V}, \bar{\mathcal{V}}, \mathcal{E}, \mathcal{C})$, where $V \in \mathcal{V}$ if V is a micro-variable, $\bar{V} \in \bar{\mathcal{V}}$ if $\bar{V}$ is a macro-variable. Further, $U \to V \in \mathcal{E}$ if U and V are micro-variables and U is a parent of V, and $S \leftrightarrow \bar{S} \in \mathcal{C}$ if $\bar{S}$ is the aggregation of S.*

The first consequence of aggregation is that the aggregated variables may no longer show a well-defined causal relation:

Example 1 *Consider the cause-effect relation $X \to Y$, with $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$. For $j = 1, 2$, set $Y_j := \alpha_j X_j$. Define $\bar{X} := \pi_X(X_1, X_2) = X_1 + X_2$, and likewise $\bar{Y} := \pi_Y(Y_1, Y_2) = Y_1 + Y_2$. Then, for $\alpha_1 \neq \alpha_2$, the effect of setting $\bar{X}$ to $\bar{x}$ on $\bar{Y}$ is ill-defined without any further specification. This example violates the consistency condition set out by Rubenstein et al. (2017). Due to space limitation, we elaborate in Appendix A.*

In reality, $\alpha_1 \neq \alpha_2$ is generic. Since the operation 'setting $\bar{X}$ to $\bar{x}$' is not well-defined *a priori*, we need to specify the distribution according to which we randomize the micro state (the 'joint manipulation' in Spirtes and Scheines (2004)):

Definition 4 (macro intervention) *A **macro-intervention** at $\bar{x}$, denoted $do(\bar{x})$, is simply a perfect intervention on the macro-variable $\bar{X}$. Given an aggregation map, $\pi : \mathcal{X} \to \bar{\mathcal{X}}$, $do(\bar{x})$ is associated with a probability measure over $\mathcal{X}$, denoted $P_{\pi,\bar{x}}^{do}(X)$, the **micro-realization**, such that $\pi(x) = \bar{x} \; \forall x \in supp(P_{\pi,\bar{x}}^{do}(X))$. When the aggregation map is clear in context, we write $P_{\bar{x}}^{do}(X)$.*

Remark *1) This simple construction represents a shift in perspective: by requiring a macro-intervention to associate with a choice of micro-realization, the actual effect of a macro-intervention can be computed naturally, addressing Challenge 2. 2) This also allows us to ask the constructive question: what are the best micro-realizations with respect to a particular goal I have on the macro-level e.g. no confounding on the macro-level?*

---

1. To understand this reversal, consider a company where the management sets guidelines of increasing total sales and the staff implement this on the micro level.
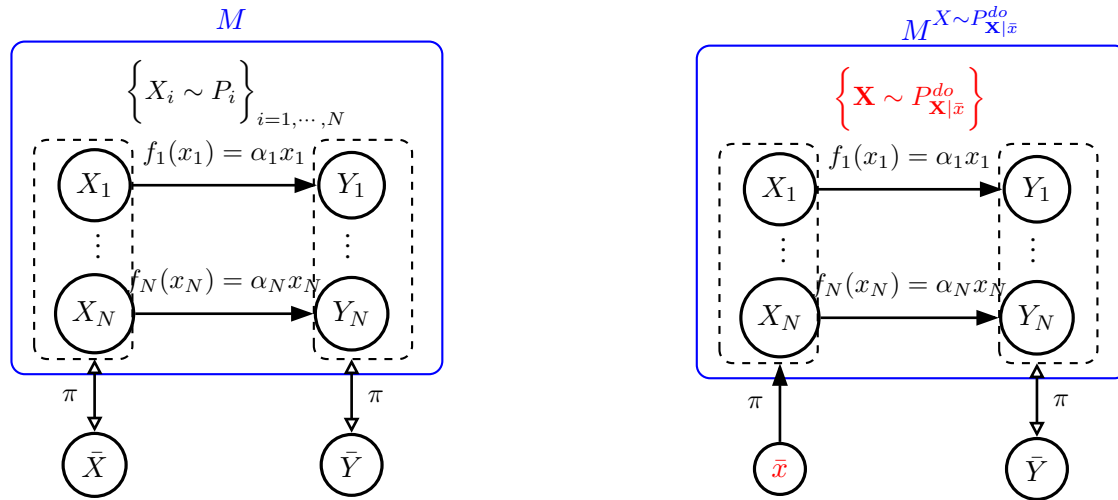
Figure 1: **Left:** The micro-variable causal model M is described by an SCM, which is shown within the blue frame. The aggregation map π is applied to the micro-variables $X_1, \cdots, X_N$ and $Y_1, \cdots, Y_N$, both contained in the dashed frame. $\bar{X}$ and $\bar{Y}$ are the macro variables which arise from the aggregation map. **Right:** A macro-intervention $P_{\pi,\bar{x}}^{do}(X)$ is shown. The intervention applied to the macro variable $\bar{X} := \bar{x}$ is realised as a (perfect) stochastic intervention on X, note that all values of $\mathbf{x}$ supported by $P_{\mathbf{X}|\bar{X}:=\bar{x}}$ give rise to $\bar{x}$.

Chalupka et al. (2016) also define a macro-level intervention via actual interventions on micro-variables, but similar to Rubenstein et al. (2017); Beckers and Halpern (2019); Massidda et al. (2023), they consider scenarios where the result is insensitive to the difference in microscopic realizations. As a further difference to their setting, we consider stochastic interventions (Correa and Bareinboim, 2020) as well as deterministic, while they only consider the latter.

The following definition formalizes reversing the causal relation between X and $\bar{X}$ with an intervention on $\bar{X}$, (see also Figure 1):

Definition 5 (macro-intervention graph) *Let $G^a = (\mathcal{V}, \bar{\mathcal{V}}, \mathcal{E}, \mathcal{C})$ be an amalgamated graph. A macro intervention $P_{\pi,\bar{x}}^{do}$ maps G^a to the macro-intervention graph, $G^{a\prime} := (\mathcal{V}, \bar{\mathcal{V}}, \mathcal{E}', \mathcal{C}')$, where $\mathcal{E}' = \mathcal{E} \bigcup \{\bar{X} \to X\} \setminus \{V \to X : V \in \mathcal{V} \setminus \{X\}\}$ and $\mathcal{C}' = \mathcal{C} \setminus \{X \leftrightarrow \bar{X}\}$.*

2.1. Confounding on the macro level

Let X and Y be two (multi-variate) micro-variables with associated macro variables $\bar{X}$ and $\bar{Y}$. Suppose that X precedes Y in causal order. Denote the observational distribution by P. Given a macro intervention at $\bar{x}$, its associated micro-realization $P_{\bar{x}}^{do}(X)$ ensures that the post-intervention quantity $P(\bar{y}|do(\bar{X} := \bar{x}))$ is well-defined: $P(\bar{y}|do(\bar{X} := \bar{x})) = \int_{\mathcal{X},\mathcal{Y}} P(\bar{y}|Y = y) P(y|do(X := x)) P_{\pi,\bar{x}}^{do}(x) dx dy$. We then define the notion of macro-confounding by comparing the observational and interventional distributions of macro variables, but first we restrict our discussion in this paper under one assumption:

Assumption 1 *We focus on the case where there is no directed causal paths among variables to be aggregated.*

This is a simplifying assumption to make discussion on defining macro-confounding easier. This allows us to say that an intervention on the aggregated variable should not affect the noise on the macro-level, which arises under a change-of-coordinate of the micro-variable space.

Definition 6 (macro confounding) *Let X and Y be two micro-variables with associated macro variables $\bar{X}$ and $\bar{Y}$ and let X precede Y in causal order. Let $\mathcal{I}$ be a set containing macro-interventions, one for each $\bar{x} \in Supp[P(\bar{X})]$. We say there is macro-confounding between macro variables $\bar{Y}$ and $\bar{X}$ if $P(\bar{Y}|\bar{X} = \bar{x}) \neq P(\bar{Y}|do(\bar{X} := \bar{x}))$ for some $\bar{x}$.*

One may question whether it is reasonable to define confounding by disagreement of interventional and observational distributions? We can justify it from the angle of a coordinate change: aggregation can be thought of as a coordinate transformation of the space of micro-variables, where one of the new coordinates is given by the aggregation map. Recall Assumption 1, there are no directed causal paths among the variables being aggregated, hence there are also no directed causal paths among the variables under the new coordinates, then an intervention on $\bar{X}$ should keep the other variables fixed. It is therefore natural to think of confounding as the correlation between $\bar{X}$ and other variables under the new coordinate system in the observational distribution, but which can be broken in the interventional. This is captured by $P(\bar{Y}|\bar{X} = \bar{x}) \neq P(\bar{Y}|do(\bar{X} := \bar{x}))$. We provide a detailed discussion of linear change-of-coordinate in Appendix C and the nonlinear case in Section 3.1.2.

Definition 7 *A **confounding-inhibiting micro-realization** of a macro-intervention has the post-interventional distribution $P(\bar{Y}|do(\bar{X} := \bar{x}))$ equal to the observational distribution $P(\bar{Y}|\bar{X} = \bar{x})$. On the other hand, a **confounding-inducing micro-realization** has $P(\bar{Y}|do(\bar{X} := \bar{x})) \neq P(\bar{Y}|\bar{X} = \bar{x})$.*

3. Confounding is ambiguous

Now we illustrate that confoundedness in the macro-level is ambiguous. We first discuss this in an unconfounded micro-model, where the macro-model may appear confounded for the simple observation that there are distinct micro-realizations to the same macro intervention. However, less obvious is the case even where the micro-model is *confounded*, there still exist cases where the micro-realizations can be chosen so that the macro-model appear unconfounded.

All proofs are found in Appendix E.

Remark *We focus mostly on linear Gaussian models, but provide preliminary results to nonlinear, non-Gaussian, and discrete cases. We leave for future work a full treatment of the general case.*

3.1. Macro-confounding in unconfounded micro-models

For simplicity of illustration, we consider a two-variable example, but the results can be readily extended to an $n-$variable linear setting.

To match the scenario in Example 1, suppose there are two identical shops, A and B, each with two products, indexed 1 and 2, with prices α_1 and α_2, respectively. The sales X_1, X_2 of the two products are assumed to be independent Gaussians with means μ_1, μ_2 and variances σ_1^2, σ_2^2. The total revenue, $\bar{Y}$ is given by $\bar{Y} = \alpha_1 X_1 + \alpha_2 X_2$. Now suppose the shop owners want to understand how total revenue changes *w.r.t.* total sales, $\bar{X} = X_1 + X_2$.

Shop A owner performs the micro-realization $P^{do}(X_1, X_2|\bar{x}) := P(X_1, X_2|\bar{x})$. Clearly this leads them to conclude that $\bar{X}, \bar{Y}$ is unconfounded, since

$$P(\bar{Y}|do(\bar{X} := \bar{x})) = P(\bar{Y}|x_1, x_2)P^{do}(x_1, x_2|\bar{x}) = P(\bar{Y}|x_1, x_2)P(x_1, x_2|\bar{x}) = P(\bar{Y}|\bar{x}).$$

Consequentially, the shop A owner concludes on the following structural equation which is what would be obtained by regressing $\bar{Y}$ on $\bar{X}$:

$$\bar{Y} = \frac{\mathrm{Cov}(\bar{Y}, \bar{X})}{\mathrm{Cov}(\bar{X}, \bar{X})}\bar{X} + N = \frac{\alpha_1\sigma_1^2 + \alpha_2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\bar{X} + N, \tag{2}$$

with $N \perp\!\!\!\perp \bar{X}$. Meanwhile, shop B owner performs an experiment also setting the total sales to $\bar{x}$, but they do this by observing how many items were sold for product 1, and then turn up (or down) the advertising to make sure they sell $\bar{x} - x_1$ items for product 2. This amounts to the macro intervention:

$$P_{\bar{x}}^{do}(X_1, X_2) = \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \bar{x} - \mu_1 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & -\sigma_1^2 \\ -\sigma_1^2 & \sigma_1^2 \end{pmatrix}\right) \tag{3}$$

Shop B owner, making this intervention, would instead conclude that the structural equation generating $\bar{Y}$ is

$$\bar{Y} = \alpha_2\bar{X} + \tilde{N}, \tag{4}$$

where $\tilde{N} = (\alpha_1 - \alpha_2)X_1$. Thus, whenever $\alpha_1 \neq \alpha_2$, we have $\tilde{N} \not\perp\!\!\!\perp \bar{X}$ under P.

Note that equations (2) and (4) generate the same observational distribution $P(\bar{Y}, \bar{X})$, but they correspond to different micro-realizations on $\bar{X}$, the former being *confounding-inhibiting* while the latter being *confounding-inducing*. This result can also be viewed from the perspective of coordinate change, which we discuss in Appendix C.

3.1.1. A SIMPLE POSITIVE RESULT

The following result provides a simple sufficient condition for which the unconfounded cause-effect relation in Figure 2, top, turns into the unconfounded cause-effect relation in Figure 2, lower:

Theorem 3.1 *Let $X \rightarrow Y$ and $\bar{X}$ and $\bar{Y}$ be macro-variables of X and Y, respectively. Then the micro-realization $P_{\bar{x}}^{do}(X) := P(X|\bar{X} := \bar{x})$ is confounding-inhibiting, i.e. the graph $\bar{X} \rightarrow \bar{Y}$ can be drawn.*

One should appreciate that this result shows an option for consistently aggregating cause-effect relations provided one is willing to accept its *context dependence*: the relation $\bar{X} \rightarrow \bar{Y}$ holds whenever one assumes that the micro-realization $P_{\bar{x}}^{do}(X)$ is just the conditional distribution of micro states X as is usual 'in the wild'. This view, however, entails that we refer to different micro-realizations in different marginal distributions of X, because conditioning macro states would then in general give rise to different distributions of micro states.
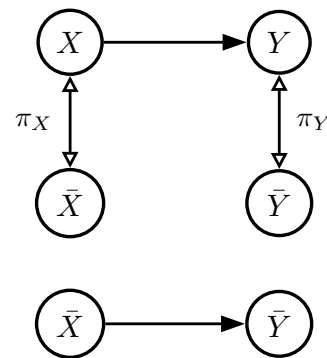
Figure 2: With an appropriate microscopic micro-realizations of interventions on $\bar{X}$, the cause-effect relation (top) remains unconfounded on the macro level (lower).

Nevertheless, adopting the distribution of micro states from the observational distribution seems like a natural definition for $P_{\bar{x}}^{do}(X)$ whenever there is no additional knowledge about the system telling us how a 'more natural' distribution would look like.[2]

Definition 8 (Natural micro-realization) *We call $P_{\bar{X}}^{do}(X) = P(X|\bar{X})$ the natural micro-realization*[3].

Definition 9 (Natural macro-intervention) *A natural macro-intervention of an aggregated variable is one that is associated with the natural micro-realization.*

Remark *Note, that the natural macro-intervention can be defined without reference to the micro-causal model. One only has to know that the underlying micro model does not confound X and Y. This addresses Challenge 3 for cases with no micro-level confounding.*

3.1.2. NON-GAUSSIAN AND NON-LINEAR GENERALIZATION

We can view $\bar{X}$ as the first coordinate after a *non-linear* coordinate transformation. We will see that even in *linear* models, noise terms that have a non-linear effect on the target may be more interpretable. Further, the most interpretable interventions may not necessarily come from *linear* coordinate changes.

To this end, assume that Y_j is given by $Y_j = f_j(X_j) + V_j, \quad j = 1, \ldots, N$, where f_j may be non-linear functions and V_j are noise terms, independent of X_j and jointly independent.

We will first show that there is an infinite continuum of options for a structural equation that writes $\bar{Y}$ in terms of $\bar{X}$ together with an appropriately constructed (formal) noise term. From these options, we will later choose one that renders the causal relation between $\bar{X}$ and $\bar{Y}$ unconfounded.

To simplify notation, we introduce the auxiliary variable $W := \sum_{j=1}^{N} f_j(X_j)$ and obtain

$$\bar{Y} = W + \bar{V}, \tag{5}$$

with $\bar{V} := \sum_{j=1}^{N} V_j$.

For our coordinate change we can restrict our attention to the two dimensional subspace spanned by $\bar{X}, W$: let $\psi : \mathbb{R}^2 \to \mathbb{R}^2$ be a bijection that leaves the first component invariant, that is $\psi(a, b) = (a, \psi_2(a, b))$. We then have $\psi^{-1}(a, b) = (a, \phi(a, b))$, where ϕ denotes the second component of ψ^{-1}.

We define an additional noise variable $M := \psi_2(\bar{X}, W)$, from which we can reconstruct W via $W = \phi(\bar{X}, M)$, and rewrite (5) as

$$\bar{Y} = \phi(\bar{X}, M) + \bar{V}, \tag{6}$$

---

2. Motivated by thermodynamics (Balian, 2007), one may alternatively want to consider the distribution with maximal entropy subject to the constraint $\bar{X} = \bar{x}$ as 'most natural'. However, entropy of continuous variables implicitly refers to a reference measure (like the volume in phase space in physics), which again introduces an ambiguity. Nevertheless, *equilibrium* thermodynamics (Adkins, 1983) provides an example where 'natural micro-realizations' exist: regardless of whether macroscopic variables like volume or pressure have been *observed* or *set* by an intervention, in both ceases the distribution of micro states is the uniform distribution (maximum entropy) in the sub-manifold of the phase space satisfying the respective macro constraint.

3. In practice, since this is stochastic, how to implement this can still be non-obvious, which is why we also introduce a deterministic analogue in Appendix B, namely shift interventions. In Appendix C.1, we point out that one can realize the same structural coefficient as the natural intervention using shift interventions.

with $\bar{V}$ being independent of $\bar{X}$ and M, but M possibly dependent of $\bar{X}$. Then, (6) can be interpreted as the SCM of a (possibly confounded) causal relation between $\bar{X}$ and $\bar{Y}$ with vector valued noise variable $(M, \bar{V})$.

To see that M can be chosen in a way that renders this relation unconfounded, define M via the conditional cumulative distribution function $M(\bar{x}, w) := P(W \leq w | \bar{X} = \bar{x})$. Whenever the conditional distribution of W, given $\bar{X}$ is continuous, M given $\bar{X}$ is uniformly distributed for all $\bar{x}$ and thus independent of $\bar{X}$. The function ϕ exists if $M(\bar{x}, \cdot)$ is invertible for all $\bar{x}$. With such a choice of M, (6) is the SCM of an unconfounded causal relation between $\bar{X}$ and $\bar{Y}$ and a micro-realization that keeps M constant is then confounding-inhibiting.

The different choices of M may differ with respect to interpretability and the one above (which renders the relation unconfounded) may not be the most interpretable one. To see this, let us revisit the example $f_j(X_j) = \alpha_j X_j$, where α_j is the price per unit and $V_j := 0$. We can then define the noise M by the average price $M := (\sum_{j=1}^{N} \alpha_j X_j) / (\sum_{j=1}^{N} X_j)$, for which we obtain the SCM $\bar{Y} = \bar{X} \cdot M$. We can then think of an intervention in which the company starts selling products to an additional country with buying patterns comparable to their existing customers. This increases the sales of all price segments by the same percentage, and thus defines an intervention that keeps the average price M constant. Generically, we will have $M \not\!\perp\!\!\!\perp \bar{X}$, and end up in the confounded scenario with an interpretable noise.

3.1.3. MACRO-INTERVENTIONS FOR WHICH WE CANNOT SPECIFY A MACRO CONFOUNDER.

So far we have shown that the macro variables appear confounded or not, depending on the specified macro interventions. As a somewhat negative result, it turns out there exist macro interventions which do not make the macro variables look unconfounded, but also do not allow for an explicit construction of a macro-confounder. Due to space limitation, we elaborate on this in Appendix D.

3.2. Macro-confounding in confounded micro-models

Even if the micro systems themselves are confounded, we may still at times get unconfounded macro-models. In 3.2.1, we first discuss the linear Gaussian setting, and describe conditions for unconfoundedness. In 3.2.2 we also present a general technical result for categorical variable models.

3.2.1. LINEAR GAUSSIAN CONFOUNDED MODELS

Assume we are given the confounded linear gaussian model

$$\bar{Y} = \alpha^T X + N, \tag{7}$$

where N is a noise variable that correlates with X but is not a descendant of X.

Let π be the aggregation that sums all elements of X. Is it possible to define $P_{\pi,\bar{x}}^{do}(X)$ in a way that is confounding inhibiting, that is, $P(\bar{Y} | do(\bar{x})) = P(\bar{Y} | \bar{x})$? The following necessary condition shows that it is not always possible, but we also derive conditions under which it is possible. Let $\Sigma_X^{do(\bar{x})}$ denote the covariance matrix of $P_{\pi,\bar{x}}^{do}(X)$. Then $P(\bar{Y} | do(\bar{x}))$ has variance

$$\text{Var}(\bar{Y} | do(\bar{x})) = \text{Var}(N) + \alpha^T \Sigma_X^{do(\bar{x})} \alpha. \tag{8}$$

Since the second term is non-negative, we can only suppress confounding if

$$\text{Var}(\bar{Y} | \bar{x}) \geq \text{Var}(N). \tag{9}$$

We can give (9) a geometric interpretation if we focus on random variables with zero mean without loss of generality and think of covariance as inner product in the Hilbert space of random variables with finite variance. We can even restrict the attention to the 3-dimensional space spanned by $\bar{X}, \alpha^T X, N$.

Accordingly, we have $\mathrm{Var}(N) = \|N\|^2$. Further, let $Q^\perp$ denote the projection onto the orthogonal complement of $\bar{X}$. Then $\mathrm{Var}(\bar{Y}|\bar{x}) = \|Q^\perp(N + \alpha^T X)\|^2$ for all $\bar{x}$, because the conditional variance is given after regressing $\bar{X}$ out and is homoscedastic since we considering joint-Gaussian variables. We can thus rewrite (9) as

$$\|Q^\perp(N + \alpha^T X)\| \geq \|N\|. \tag{10}$$

To see that (9) is also sufficient to enable confounding inhibiting interventions, we will briefly check that $\alpha^T \Sigma_X^{do(\bar{x})} \alpha$ can be made arbitrarily large, and then it follows (9) is sufficient to guarantee existence of a Gaussian micro-realization that is confounding-inhibiting. Whenever the two vectors α and $\mathbf{1} = (1, \ldots, 1)^T$ are linearly independent (otherwise the problem is anyway trivial), this is achieved by choosing $\Sigma_X^{do(\bar{x})}$ with $\mathbf{1}$ being an eigenvector with eigenvalue 0 and other eigenvalues large enough so that $\alpha^T \Sigma_X^{do(\bar{x})} \alpha = \mathrm{Var}(\bar{Y}|\bar{x}) - \mathrm{Var}(N)$.

We now need to show that we can define $do(\bar{x})$ in a way that ensures that conditional expectations also match:

$$E[\bar{Y}|do(\bar{x})] = E[\bar{Y}|\bar{x}]. \tag{11}$$

Since the intervention cannot affect N by assumption, (7) implies $E[\bar{Y}|do(\bar{x})] = \sum_j \alpha_j E[X_j|do(\bar{x})] + E[N]$.

In defining our intervention, we can freely choose each $E[X_j|do(\bar{x})]$ with the only constraint $\sum_j E[X_j|do(\bar{x})] = \bar{x}$. Whenever α and $(1, \ldots, 1)$ are linearly independent, the term $\sum_j \alpha_j E[X_j|do(\bar{x})]$ can be made to achieve *any* real value, hence we can certainly ensure (11).

Finally, for which values of α is (10) (or equiv., (9)) satisfied? When $\alpha \propto \mathbf{1}$, it is clearly not satisfied: since $Q^\perp \bar{X} = 0$ and $\|Q^\perp N\| < \|N\|$ as $\bar{X}$ and N are not perpendicular under the covariance inner product. But as soon as α and $\mathbf{1}$ are linearly independent, the RHS of (8) can be made arbitrarily large by scaling up α while having it always point in the same direction. Thus, it is possible to tune α as a hyperparameter of the system to ensure that confounding inhibiting interventions exist.

Figure 3: Confounded cause-effect relation between X and Y. **Left**: G^a. The amalgamated graph before macro intervention on $\bar{X}$. **Right**: $G^{a'}$. The amalgamated graph after macro intervention on $\bar{X}$.

Interpretation. For the relation between sold items X_i of each product i with price α_i and $Y_i = \alpha_i X_i$ its revenue, let Z_i be the number of extra products sold on promotion with reduced price γ_i. Assume that if the customer buys one item on reduced price, then they will also buy one on full price, giving rise to the micro-level SCM $X_i := Z_i$ and $Y_i := \alpha_i X_i + \gamma_i Z_i$. Here, Z_i are micro confounders of the cause-effect
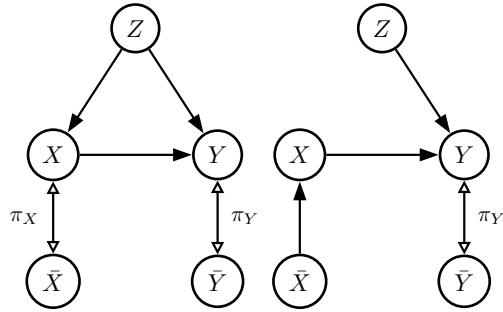
pair $X_i \to Y_i$ and we obtain an aggregated confounder $N := \sum_i \gamma_i Z_i$, which may disappear for appropriate interventions on $\bar{X}$, provided the prices α_i are 'different enough' from each other.

The above formulation asserts that the shop owner selling the items can make a confounding-inhibiting micro-realization on the total number of items sold on full price ($\bar{X}$) when full prices $\boldsymbol{\alpha}$ contain enough heterogeneity wrt to the reduced prices $\boldsymbol{\gamma}$ i.e. When all prices are the same, there is not enough heterogeneity in the micro-model to blur out the confounding. On the other hand, if there is enough heterogeneity in the prices, then the shop owner may find a macro-intervention on the items sold on full prices, for which they can infer the causal effect of, by directly regressing total revenue on total items sold on full price.

3.2.2. DISCRETE CONFOUNDED MICRO MODELS

In the case of nonlinear and non-Gaussian models, we provide at least a result in the case with discrete variables. Our result says that even if the system remains confounded after aggregation in the effect variable, the system may still regain unconfoundedness after aggregation of the cause variables. The proof is in Appendix E.2.

Theorem 3.2 *Let X, Y, Z be categorical variables with $|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}| < \infty$. Let G be the causal graph with $X \to Y$, $X \leftarrow Z \to Y$, and let $\pi_X : \mathcal{X} \longrightarrow \bar{\mathcal{X}}$ and $\pi_Y : \mathcal{Y} \longrightarrow \bar{\mathcal{Y}}$ denote aggregation maps. Let G^a be the amalgamated graph of G, π_X and π_Y. Further, let $G^{a\prime}$ be the post-intervention graph for a macro-intervention on $\bar{X}$. See Figure 3 Let P denote probability distributions satisfying the conditional independence structure given by G^a, and P^{do} denote probability distributions satisfying the independence structure given by $G^{a\prime}$. Suppose $|\bar{\mathcal{Y}}| < \min(|\mathcal{X}|, |\mathcal{Z}|)$, $|\bar{\mathcal{X}}| < |\mathcal{X}|$. Then for any coarsening maps π_X and π_Y, and any family of macro interventions on $\bar{X}$: $\{P^{do}_{\pi_X, \bar{x}} \mid \bar{x} \in \bar{\mathcal{X}},\ P^{do}_{\pi_X, \bar{x}} \neq P_{X|\bar{X}=\bar{x}}\}$ there exist at least one P such that*

1. *$P(Y|X) \not\equiv P^{do}(Y|X)$, $P(\bar{Y}|X) \not\equiv P^{do}(\bar{Y}|X)$. i.e. (X,Y) confounded, $(X, \bar{Y})$ confounded.*

2. *$P(\bar{Y}|\bar{X}=\bar{x}) = P^{do}(\bar{Y}|\bar{X}=\bar{x})$ for all $\bar{x} \in \bar{\mathcal{X}}$. i.e. $\{(\bar{x}, P^{do}_{X|\bar{X}}) \mid \bar{x} \in \bar{\mathcal{X}}\}$ are confounding-inhibiting micro-realizations.*

The conditions $|\bar{\mathcal{Y}}| < \min(|\mathcal{X}|, |\mathcal{Z}|)$ and $|\bar{\mathcal{X}}| < |\mathcal{X}|$ are sufficient conditions which ensure that there is strong-enough aggregation. The theorem thus states that we could find joint distributions over $X, Y, \bar{X}, \bar{Y}$ such that even if both (X,Y) and $(X, \bar{Y})$ are confounded, the confounding can be 'cancelled out' for $(\bar{X}, \bar{Y})$ provided there is enough aggregation.

4. Multi-variate models

In this section we want to follow up on the positive result of Theorem 3.1 and discuss generalizations and their challenges.

4.1. Linear chain

Given the micro model in Figure 4, top, we are now facing two challenges if we wish to aggre-
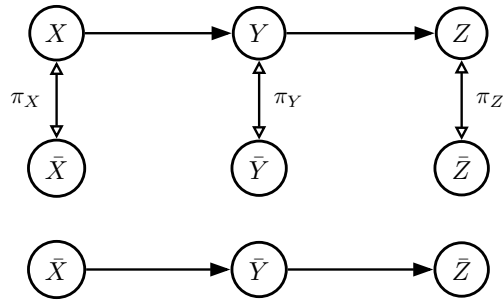

Figure 4: The *natural micro-realization* of macro interventions from Section 3.1.1 enables coarse graining the chain (top) to the chain (lower) whenever $\bar{X} \perp\!\!\!\perp \bar{Z} | \bar{Y}$.

gate it to the structure on the bottom. First, the macro variable $\bar{Y}$ will not necessarily block the

influence of $\bar{X}$ on $\bar{Z}$ due to information transmission through the micro state Y. Second, the generalization of *natural micro-realization* raises the following ambiguity for $do(\bar{y})$: to someone who focuses on the cause-effect relation $Y \to Z$ only, it should be $P(Y|\bar{y})$, aligned with Section 3.1.1. However, after seeing $\bar{X}$ one may want to consider $P(Y|\bar{y}, \bar{x})$ more natural. This appears to be the right choice at least if one restricts the analysis to the sub-population with fixed $\bar{x}$. Remarkably, these two questions are related:

Lemma 10 (irrelevance of cause of causes) *Given the causal chain in Figure 4, then setting $do(\bar{y}) :=$ $P(Y|\bar{y})$ results in the same downstream effect on $\bar{Z}$ as implementing it according to $P(Y|\bar{y}, \bar{x})$ if and only if*

$$\bar{X} \perp\!\!\!\perp \bar{Z} \,|\, \bar{Y}. \tag{12}$$

If (12) does not hold, the chain in Figure 4, lower, is not a valid aggregation and requires an additional arrow $\bar{X} \to \bar{Z}$ instead.[4] Whenever (12) holds, the cause-effect relation $\bar{Y} \to \bar{Z}$ is valid with respect to both micro-realizations $P(Y|\bar{y}, \bar{x})$ and $P(Y|\bar{y})$, and they result in same effect on $\bar{Z}$. Since the natural micro-realization of $do(\bar{x})$ has no ambiguity and renders the relation to $\bar{Y}$ and $\bar{Z}$ unconfounded, the chain in Figure 4, lower, captures all causal relations *w.r.t.* natural micro-realizations. Conveniently, (12) is a purely statistical criterion about macro variables only.

4.2. Backdoor adjustments

Assume now we have the DAG in Figure 5, left.

Under which conditions can we define micro-realizations of macro interventions with respect to which the system behaves like the DAG on the right of Figure 5? We already know that $P(X|\bar{x})$ yields effects on $\bar{Y}, \bar{Z}$ that align with the observational conditionals $P(\bar{Y}|\bar{X})$ and $P(\bar{Z}|\bar{X})$, respectively. The questionable part is the effect of interventions on $\bar{Y}$. We need to define them in a way that ensures

$$P(\bar{Z}|do(\bar{y}), \bar{x}) = P(\bar{Z}|\bar{y}, \bar{x}), \tag{13}$$

that is, the backdoor adjustment formula with $\bar{X}$ as adjustment variable (Rule 2, Theorem 3.4.1 in Pearl (2009)). We find the following sufficient condition:

Lemma 11 (macro backdoor adjustment) *Given the DAG Figure 5, left, let $\bar{Z}_y$ denote the random variable $\bar{Z}$ after adjusting Y to some fixed y in the SCM for $\bar{Z}$. If*

$$\bar{Z}_y \perp\!\!\!\perp Y \,|\, \bar{X}, \tag{14}$$

then (13) holds if $do(\bar{y})$ is defined by randomizing Y according to $P(Y|\bar{y}, \bar{x})$.
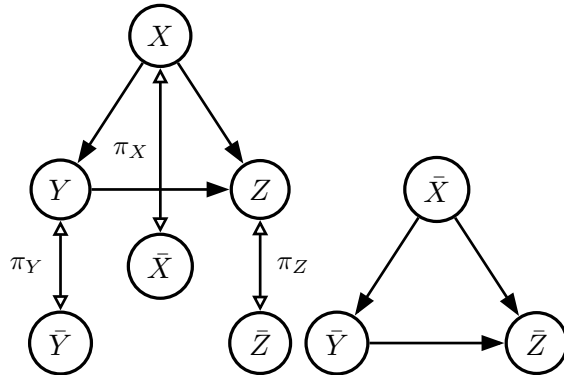

Figure 5: Left: complete DAG with micro variables X, Y, Z, together with its aggregations $\bar{X}, \bar{Y}, \bar{Z}$. Right: The aggregated DAG which we would like to give a causal semantics by introducing appropriate interventions, if possible.

---

4. In the context of dynamical processes, this question translates to the difficult question of which macro variables are required to describe the relevant part of the history, see e.g. Crutchfield and Shalizi (1999).

Lemma 11 states that we can consider the DAG in Figure 5, right, as a valid aggregation of the one on the left when we make the macro-intervention $P(Y|\bar{y}, \bar{x})$, provided that $\bar{X}$ blocks the backdoor path between Y and $\bar{Z}$. Condition (14) is strong, since it refers to the micro state of Y; finding conditions that are easier to handle is left to the future. Nevertheless, it is worth mentioning that the proof of Lemma 11 straightforwardly generalizes to back-door adjustments in arbitrary DAGs. We first state Definition 3.3.1 in (Pearl, 2009).

Definition 12 (back-door criterion) *A set B of variables satisfies the back-door criterion relative to the ordered pair (X_i, X_j) if (i) no node in B is a descendant of X_i and (ii) B blocks every path between X_i and X_j that contains an arrow into X_i.*

Theorem 4.1 (Generalized backdoor adjustment) *Let G be the causal DAG for $X_1, \ldots, X_n$ and $P(\bar{X}_1, \ldots, \bar{X}_n)$ be Markovian relative to some DAG $\bar{G}$ that contains $\bar{X}_i \to \bar{X}_j$ if $X_i \to X_j$ is an arrow in G, but possibly also additional arrows. Let the set B of macro variables satisfy the back-door criterion with respect to the pair $(\bar{X}_i, \bar{X}_j)$ in $\bar{G}$ and*

$$(\bar{X}_j)_{x_i} \perp\!\!\!\perp X_i \,|B, \tag{15}$$

where $(\bar{X}_j)_{x_i}$ denotes $\bar{X}_j$ after adjusting X_i to x_i. Then we have $P(\bar{X}_j|do(\bar{x}_i), b) = P(\bar{X}_j|\bar{x}_i, b)$ with respect to the micro-realization $P(X_i|\bar{x}_i, b)$.

Whenever there are different adjustment sets B, B' with respect to which condition (15) holds, the micro-realizations $P(X_i|\bar{x}_i, b)$ and $P(X_i|\bar{x}_i, b')$ result in the same effect on $\bar{X}_j$. This is because in any DAG, interventional distributions computed from different back-door adjustments coincide.

5. Example application scenarios

Here we detail two scenarios where our framework can be used in practice.

Unconfounded micro-model. A large retail company has millions of products for sale. The micro variables are sales of products in some same category $X_1, \ldots, X_K$, for example, X_i is the number of bottles sold for the ith type of shampoo, and product revenues $Y_1, \ldots, Y_K$. The value of K is large, i.e. in the thousands. The micro-variable structural equations are $Y_i = \alpha_i X_i + N_i$, $X_i = \mathcal{N}(\mu_i, \sigma_i)$, and $N_i \perp X_i$. For the management, a macro model is presented, where the macro variables are $\bar{X}$ and $\bar{Y}$. The management would like to enable easy causal inference at the macro level by reading off the causal effect of changing $\bar{X}$ on $\bar{Y}$ from the observational distribution $P(\bar{Y}|\bar{X})$. This instructs the natural micro-realization, that is $P_\pi^{do}(X|\bar{X} = \bar{x}) := P(X|\bar{X} = \bar{x})$ when implementing the macro intervention $\bar{X} := \bar{x}$. Note, that the natural micro realization does not reference the micro level structural equations. In practice, the natural implementation here is a Multivariate Gaussian density whose parameters are functions of $\bar{x}$, so the company can sample from the density and set $X_i, \ldots, X_K$ to the sampled value. The caveat is that in practice this procedure can become costly.

Confounded micro-model. Now imagine the same micro-level model, but this time X_i is not independent of N_i. There are two ways to proceed. (i) The management still wants to read off causal effect directly from the observational distribution $P(\bar{Y}|\bar{X})$. Then the company needs to implement a confounding inhibiting micro realization, for which we described the derivation and sufficient conditions in section 3.2.1. Note that the sufficient condition, i.e. having enough variability in the prices α_i now depends on the micro structural equations. (ii) If we can define $\bar{N}$ as a function of $X_1, \ldots, X_K, N_1, \ldots, N_K$ such that equation (14) is satisfied, that is $\bar{Y}_x \perp X|\bar{N}$ then the company can also perform a macro-backdoor adjustment $P(\bar{Y}|do(\bar{X})) = \sum_{\bar{N}} P(\bar{Y}|\bar{X}, \bar{N})P(\bar{N})$.

6. Conclusion

Aggregated variables are the rule rather than the exception in causal models of processes in everyday life. We highlight three challenges for causal aggregation which are not sufficiently addressed by existing work: 1) causal aggregation induces ambiguity, 2) bridging macro- and micro-interventions, and 3) the micro-graph may not be known. This paper intensifies Challenge 1 and shows that causal aggregation entails ambiguity not only with respect to the quantitative effect, but even with respect to the causal structure, namely whether a cause-effect relation is confounded or not. On the positive side, this paper addresses Challenges 3 by introducing 'natural micro-realizations', defining which does not require knowledge of the causal graph. Moreover, in introducing 'natural macro-interventions' and 'confounding-inhibiting (inducing) micro-realizations', we naturally address Challenge 2 and open up the way to ask questions of 'What are the best micro-realizations for a given macro-intervention under a given goal such as no macro-confounding?'.

Acknowledgments

References

C. J. Adkins. *Equilibrium Thermodynamics*. Cambridge University Press, 3 edition, 1983.

Tara V. Anand, Adele H. Ribeiro, Jin Tian, and Elias Bareinboim. Causal effect identification in cluster dags. AAAI Press, 2023.

Roger Balian. *From microphysics to macrophysics*, volume 1. Springer, 2007.

Sander Beckers and Joseph Y. Halpern. Abstracting causal models. In *AAAI*, 2019.

Sander Beckers, Frederick Eberhardt, and Joseph Y. Halpern. Approximate causal abstraction. *UAI*, 2019.

Sander Beckers, Joseph Halpern, and Christopher Hitchcock. Causal models with constraints. In *Proceedings of the Second Conference on Causal Learning and Reasoning*, 2023.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: a review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8): 1798—1828, August 2013.

Tineke Blom, Stephan Bongers, and Joris M. Mooij. Beyond structural causal models: Causal constraints models. In *UAI*, 2019.

Krzysztof Chalupka, Frederick Eberhardt, and Pietro Perona. Multi-level cause-effect systems. In *AISTATS*, 2016.

Juan Correa and Elias Bareinboim. A calculus for stochastic interventions:causal effect identification and surrogate experiments. In *AAAI*, volume 34, pages 10093–10100, Apr. 2020.

James P. Crutchfield and Cosma Rohilla Shalizi. Thermodynamic depth of causal states: Objective complexity via minimal representations. *Phys. Rev. E*, 59:275–283, Jan 1999.

Dominik Janzing and Sergio Garrido Mejia. Phenomenological causality. 11 2022. doi: 10.48550/arXiv.2211.09024.

Riccardo Massidda, Atticus Geiger, Thomas Icard, and Davide Bacciu. Causal abstraction with soft interventions. In *Proceedings of the Second Conference on Causal Learning and Reasoning*, 2023.

Edward Miguel, Shanker Satyanath, and Ernest John Sergenti. Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, 112:725 – 753, 2004.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009.

Jonas Peters, Dominik Janzing, and Bernhard Schlkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press, 2017. ISBN 0262037319.

Eigil Fjeldgren Rischel and Sebastian Weichwald. Compositional abstraction error and a category of causal models. In *UAI*, 2021.

Dominik Rothenhäusler, Christina Heinze, Jonas Peters, and Nicolai Meinshausen. Backshift: Learning causal cyclic graphs from unknown shift interventions. In *NeurIPS*, 2015.

Paul K. Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M. Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. In *UAI*, 2017.

Numair Sani, Jaron Lee, and Ilya Shpitser. Identification and estimation of causal effects defined by shift interventions. In *UAI*, 2020.

B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio. Towards causal representation learning. *Proceedings of the IEEE*, 109(5):1–23, 2021.

Peter Spirtes and Richard Scheines. Causal inference of ambiguous manipulations. 71:833–845, 2004.

Fabio Massimo Zennaro, Paolo Turrini, and Theodoros Damoulas. Quantifying consistency and information loss for causal abstraction learning. In *IJCAI*, 2023.

Appendix A. Example from Rubenstein et al. (2017)

Example 1 *Let $M = (\mathbf{S}, P_\mathbf{N})$ be a micro causal model where*

$$\mathbf{S} = \{X_1 := U_1, \ X_2 := U_2, \ Y_1 := \alpha_1 X_1, \ Y_2 := \alpha_2 X_2\}$$

where $\alpha_1 \neq \alpha_2$, $\mathcal{U}_1 = \{0,1\}$, $\mathcal{U}_2 = \{0,1\}$, $supp[P_{U_1}] = supp[P_{U_2}] = \{0,1\}$. Now, consider a set of interventions that we are interested in on the micro causal model.

$$\mathcal{I} = \{i_1 = do(X_1 := 1, X_2 := 0),$$
$$i_2 = do(X_1 := 0, X_2 := 1)\}.$$

as well as aggregation maps $\bar{X} = \pi_X((X_1, X_2)) = X_1 + X_2$ and $\bar{Y} = \pi_Y((Y_1, Y_2)) = Y_1 + Y_2$.

Now applying i_1 to the system results in $X_1 = 1$, $X_2 = 0$, $Y_1 = \alpha_1$, $Y_2 = 0$, and thus $\bar{X} = 1$ and $\bar{Y} = \alpha_1$. Meanwhile, applying i_2 to the system results in $X_1 = 0$, $X_2 = 1$, $Y_1 = 0$, $Y_2 = \alpha_2$, and thus $\bar{X} = 1$ and $\bar{Y} = \alpha_2$. Thus, there is no structural assignment $f : \bar{\mathcal{X}} \to \bar{\mathcal{Y}}$ s.t.

$$\pi_\# P^{do(i_k)}(X_1, X_2, Y_1, Y_2) = P^{do(\bar{X} := 1)}(\bar{X}, \bar{Y})$$

where $k = 1, 2$.

Therefore, it is not possible to define a sensible causal model on the macro variables: since X_1, X_2 precede Y_1, Y_2 in causal order, $\bar{X}$ should also precede $\bar{Y}$ if there is a 'macro causal model', yet there is no way to define a structural equation from $\bar{X}$ to $\bar{Y}$ that makes the 'macro causal model' consistent with the micro model in the sense that the pushforward measure of interventions on the micro model is always equal to the interventional distribution on the macro model given by the corresponding intervention. Here, corresponding intervention is given by the aggregation map. The above intuition can be formalized - (Rubenstein et al., 2017) provides a formal definition of consistency, which is violated by our example exactly for the reason we laid out. In fact, only when $\alpha_1 = \alpha_2$ are we able to obtain a consistent macro causal model under the definition of theirs.

Appendix B. Digression on Shift Interventions

The natural micro-realization is a stochastic intervention on the micro-variables. How to implement this is not immediately obvious. Thus, here we suggest shift interventions, which change the value of micro-variables *relative* to its current state. One choice is to consider the simple shift-intervention $X \mapsto X + \delta$ with constant δ, which is a special case of general shift-interventions $X \mapsto f(X)$ (Rothenhäusler et al., 2015; Sani et al., 2020). In the standard setting (i.e. without dealing with aggregation), a shift intervention is implemented as the practitioner simply observe whichever treatment the subject is about to receive, and then add a constant on top of it. Crucially, this is a deterministic action. We now elaborate on its relations with atomic interventions.

When treatments take real values, atomic interventions can be generated from shift-interventions:

$$\mathbb{P}(Y|do(X := x))$$
$$= \int_{x' \in \mathbb{R}} \mathbb{P}(Y|X = x', do\left(X := X + (x - x')\right))$$
$$\cdot \mathbb{P}(X = x')dx' \tag{16}$$
$$= \int_{x' \in \mathbb{R}} \mathbb{P}(Y|X = x', do(X := x)) \cdot \mathbb{P}(X = x')dx' \tag{17}$$

and vice versa:

$$\mathbb{P}(Y|do(X := X + \delta))$$
$$= \int_{x' \in \mathbb{R}} \mathbb{P}(Y|X = x', do\left(X := x' + \delta\right)))$$
$$\cdot \mathbb{P}(X = x')dx' \tag{18}$$

In other words, we can always generate the effect of an atomic intervention from shift interventions, and vice versa: imagine a practitioner observing a particular treatment-assignment group (x'), they then make the appropriate shift-intervention $(x - x')$ to change the treatment to x; they do this for every observed x'-treatment-assignment group. Then on average, this is the same as doing an atomic intervention to x for the whole group.

We can come up with an equivalent notion of no confounding using shift-interventions. Therefore, we may reason safely with shift-interventions and assured that whenever we conclude unconfoundedness under shift-interventions, we will also conclude unconfoundedness under atomic interventions.

Definition 13 (Unconfoundedness under δ-shift-interventions for 1-d Linear Gaussian models)
Suppose the causal order is $X \to Y$. Define the counterfactual variable after a δ-shift intervention as $X^\delta := X + \delta$. X is unconfounded with Y if

$$\forall \delta \in \mathbb{R}, \forall x \in \mathbb{R}$$
$$\mathbb{P}(Y|X^\delta = x) = \mathbb{P}(Y|X = x) \tag{19}$$

The intuition of this is the following: imagine patient A and patient B would have received treatment 1.0 and treatment 1.1 without intervention. The shift-intervention would add $+0.1$ to all treatments which the patients would have received. So, after the intervention, patient A and patient B would receive treatment 1.1 and treatment 1.2. Note that patient *A after* the intervention and patient B *before* the intervention receive the same treatment (1.1). If unconfoundedness holds, then patient B before intervention and patient A after intervention would react the same way.

We now prove the equivalence of this definition to the one using atomic interventions, in the case of linear structural equations.

Lemma 4 *Define the post-shift-intervention variable as $X^\delta := X + \delta$. For a linear structural equation model, $Y := aX + N$ where a is a constant coefficient, the following two statements are equivalent:*

1. $\forall x \in \mathcal{X} = \mathbb{R}, \ P(Y|do(X := x)) = P(Y|X = x)$.

2. $\forall \delta \in \mathbb{R}, \ \forall x \in \mathbb{R}, \ P(Y|X^\delta = x) = P(Y|X = x)$,

Proof $(2 \implies 1.)$ By definition,

$$P(Y|X^\delta = x) = P(Y|X = x - \delta, do(X := x)) \ \ \forall \delta \tag{20}$$

By 2,

$$P(Y|X = x', do(X := x)) = P(Y|X = x)\forall x' \tag{21}$$

Therefore,

$$P(Y|do(X := x)) = \int_{\mathcal{X}} P(Y|X = x', do(X := x)) \underbrace{P(X = x'|do(X = x))}_{=P(X=x')} dx' \tag{22}$$

$$= \int_{\mathcal{X}} P(Y|X = x)P(X = x')dx' \qquad \text{by (21)} \tag{23}$$

$$= P(Y|X = x) \tag{24}$$

$(1 \implies 2.)$ By 1,

$$P(Y|do(X := x)) = \int_{\mathcal{X}} P(Y|X = x', do(X := x))P(X = x'|do(X := x))dx' \tag{25}$$

$$= \int_{\mathcal{X}} P(Y|X = x', do(X := x))P(X = x')dx' \tag{26}$$

$$= P(Y|X = x) \ \forall x \tag{27}$$

Since $Y = aX + N$, we have

$$(26) = \int_{\mathcal{X}} P(ax + N|X = x')P(X = x')dx' \tag{28}$$

$$= P(ax + N) \ \text{(by basic rules of probability)} \tag{29}$$

$$= P(ax + N|X = x) = (27) \ \forall x \tag{30}$$

This means that $N \perp\!\!\!\perp X$, and therefore 1 holds. ∎

Appendix C. A perspective from linear change of coordinates

The above example can also be formalized by thinking of the vector of micro variables $X = (X_1, \cdots, X_N)$ as a coordinate system and the aggregation as a result of a linear change of coordinates. Let $H : \mathbb{R}^N \to \mathbb{R}^N$ be a bijective linear map (i.e. the change-of-coordinate map). Further, requiring the first row of H to be ones would give us the sum aggregation:

$$H_{1,:} = (1, 1, \ldots, 1) \tag{31}$$

Then, trivially it follows that

$$\bar{Y} := \boldsymbol{\alpha}^\top X = \boldsymbol{\alpha}^\top H^{-1} H X = \underbrace{\sum_{i=1}^{N} \alpha_i H_{i,1}^{-1} \bar{X}}_{\beta} + \underbrace{\sum_{i,k=1}^{N} \sum_{j=2}^{N} \alpha_i H_{i,j}^{-1} H_{j,k} X_k}_{\bar{U}_\beta}$$

Therefore, after a linear change of coordinates with $\bar{X}$ being part of the new coordinate system, $\bar{Y}$ still admits a linear structural equation under the new coordinates. We can view this as a usual structural equation of $\bar{Y}$, since all we did was changing the coordinates. Now, an intervention on $\bar{X}$ amounts to changing $\bar{X}$ while not affecting remaining coordinates $\sum_{k=1}^{N} H_{2:,k} X_k$, and hence the

noise term, $\bar{U}_\beta$, which is a linear combination thereof. This translates the ambiguity of $do(\bar{x})$ into specifying the basis vectors corresponding to the *remaining* coordinates.

One may ask, since H is constrained - it must satisfy (31) - it needs to be shown that there exist H so that a given β can be obtained. With a little bit more effort, we can also show that for any $\mathbf{\Delta}$ such that $\sum_i \Delta_i = 1$, there also exist H such that $\mathbf{\Delta} = H_{:,1}^{-1}$. Indeed, we check this in the following lemma.

Lemma 14 *For a given $\boldsymbol{\alpha} \in \mathbb{R}^N$, $\boldsymbol{\alpha} \not\propto (1, \cdots, 1)$ and $c \in \mathbb{R}$, there exists an invertible $N \times N$ matrix $H \in \mathbb{R}^{N \times N}$ such that $H_{1,i} = 1 \; \forall i = 1, \cdots, N$ and $\sum_i \alpha_i H_{i,1}^{-1} = c$. Moreover, for any $\mathbf{\Delta}$ such that $\sum_i \Delta_i = 1$, there also exist H such that $\mathbf{\Delta} = H_{:,1}^{-1}$.*

The proof is in Appendix E. Lemma 14 shows that for any β, there exist at least one corresponding change-of-coordinate matrix H. This is remarkable: When $\beta = \frac{\text{Cov}(\bar{Y}\bar{X})}{\text{Cov}(\bar{X},\bar{X})}$, we know from regression that $\bar{X}$ and $\bar{Y}$ look unconfounded. On the contrary, note that $\bar{U}_\beta = \boldsymbol{\alpha}^\top X - \beta\bar{X}$. Therefore, as $\beta \to \infty$, $|\text{Cov}(\bar{X}, \bar{U}_\beta)| \to \infty$. This means that depending on the change-of-coordinate matrix chosen, $\bar{X}$ can look either unconfounded or *arbitrarily strongly confounded with $\bar{Y}$*.

Relating the ambiguity of macro interventions to the ambiguity of the coordinate systems links this work also to causal representation learning, which deals with causal modeling in scenarios where the variables are not given a priori (Bengio et al., 2013; Schölkopf et al., 2021).

C.1. Shift interventions under a change-of-coordinate.

Clearly, the change-of-coordinate perspective also implies, as a consequence, how to distribute the shift in $\bar{X}$ to shifts in the micro-variables $X_1, \cdots, X_N$. From (32), doing $\bar{X} \mapsto \bar{X} + 1$ amounts to doing $X \mapsto X + H_{:,1}^{-1}$:

$$\bar{Y}(\bar{X} + 1) = \underbrace{\boldsymbol{\alpha}^\top H_{:,1}^{-1}\bar{X} + U_\beta}_{\boldsymbol{\alpha}^\top X} + \boldsymbol{\alpha}^\top H_{:,1}^{-1} = \boldsymbol{\alpha}^\top (X + H_{:,1}^{-1}) \tag{32}$$

i.e. shifting by 1 on $\bar{X}$ and shifting X by $H_{:,1}^{-1}$ on the micro-variables are the same thing.

This makes calculating β easy if you have already decided how you want to shift X: by Lemma 14, $H_{:,1}^{-1}$ can be any vector $\mathbf{\Delta}$ whose elements sum to 1. Therefore, just make sure that the coefficients of Δ sum up to 1, and the corresponding structural coefficient reads $\beta = \boldsymbol{\alpha}^\top \Delta$. Importantly, we realize that by setting $\Delta = \frac{1}{\sum_n \text{Var}[X_n]}(\text{Var}[X_1], \cdots, \text{Var}[X_N])$, we actually obtain the same β as that obtained by the natural intervention.

Appendix D. There are macro-interventions for which we cannot specify a macro confounder.

We have shown that depending on the specified macro interventions, the macro variables appear confounded or not. As we will show next, there also exist macro interventions which do not allow for any model on the macro variables that explains both the observational and interventional distribution. Suppose in the scenario of Section 3 we have a third shop owner who made the deterministic intervention

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \Big| \bar{X} \sim \mathcal{N}\left(\begin{pmatrix} \bar{X} \\ 0 \end{pmatrix}, \mathbf{0}\right) \tag{33}$$

This would induce a deterministic post-intervention relationship between $\bar{\mathcal{X}}$ and $\bar{\mathcal{Y}}$:

$$\bar{Y} = \alpha_1 \bar{X} \tag{34}$$

Had there been a structural causal model from $\bar{X}$ to $\bar{Y}$ under this intervention, it would have the form:

$$\bar{Y} = f(\bar{X}, N) \tag{35}$$

but since intervening on $\bar{X}$ to set $\bar{Y}$ to a deterministic value, f must be constant in N for every $\bar{X} = \bar{x}$, so wlog we can rewrite the structural equation as

$$\bar{Y} = g(\bar{X}) \tag{36}$$

But this implies that the observational random variable, $\bar{Y}|\bar{X}$, is also deterministic, which contradicts our hypothesis wherever $\alpha_1 \neq \alpha_2$. Thus, there may be macro-interventions for which we cannot write down a macro causal model which is possibly confounded. We leave it for future work to further investigate the range of macro-interventions which admit macro-confounding variables.

Appendix E. Proofs

E.1. Section 3.1

Proof of Theorem 3.1. To keep notation simple, we will consider discrete variables but the generalization to continuous variables is obvious.

$P(\bar{Y}|do(\bar{x})) = \sum_{x,y} P(\bar{Y}|y)P(y|x)P(x|\bar{x}) = \sum_x P(\bar{Y}|x)P(x|\bar{x}) = P(\bar{Y}|\bar{x})$. ∎

E.2. Section 3.2. Micro-confounding - discrete case

When in the discrete setting, we use the standard way to construct probability spaces i.e. take the entire set as sample space and its power set as the σ-algebra.

We outline here the moral reason that the statement should be true.

For the observational and interventional distributions of $\bar{Y}|\bar{X}$ to be the same, i.e. unconfoundedness in the macro variables, we can consider the decomposition of $P_{\bar{Y}|\bar{X}}$ and $P_{\bar{Y}|\bar{X}}^{do}$. Since the macro interventions, $(\bar{x}, P_{X|\bar{x}})$ for various values of $\bar{x}$, only impact the generative process for X and $\bar{X}$, we can observe that all other generative mechanisms are invariant before and after the intervention. Writing down the decomposition, we get

$$P_{\bar{Y}|\bar{X}}^{do}(y) = \sum_{i,j} P_{\bar{Y}|x_i,z_j}(y) P_{X|\bar{X}}^{do}(x_i) P_Z(z_j) \tag{37}$$

$$P_{\bar{Y}|\bar{X}}(y) = \sum_{i,j} P_{\bar{Y}|x_i,z_j}(y) P_{X,Z|\bar{X}}(x_i, z_j) \tag{38}$$

Unconfoundedness between $\bar{X}$ and $\bar{Y}$ thus amounts to saying that the right-hand-sides of (37) and (38) are the same for all values of y. (37) and (38) can be written as vectorised equations, and equality of the right-hand-side amounts to saying that, roughly speaking, the difference between $P_{X|\bar{X}}^{do} \otimes P_Z$ and $P_{X,Z|\bar{X}}$ lie in the null space of $P_{\bar{Y}|X,Z}$. Why should this be true? Morally, this is because the null space of $P_{\bar{Y}|X,Z}$ is large due to the coarsening, thus possible to contain the said

difference sometimes. Additional technical conditions need to be satisfied in the theorem, but these can likewise be satisfied by constructing $P_{\bar{Y}|X,Z}$ with the correct null space.

Proof of Theorem 3.2. The set of probability measures with conditional dependence structure encoded by G^a is

$$\mathcal{P}^{obs} := \left\{ P \ s.t. \ P(X, Z, \bar{X}, \bar{Y}) \equiv P(Z)P(X|Z)P(\bar{X}|X)P(\bar{Y}|X, Z) \right\} \tag{39}$$

Meanwhile, $\mathcal{P}^{do}$ denotes the set of probability measures P^{do} on the same sample space and σ-algebra, but compatible with performing a macro-intervention on $\bar{X}$:

$$\mathcal{P}^{do} := \left\{ P^{do} \ s.t. \ P^{do}(X, Z, \bar{X}, \bar{Y}) \equiv P^{do}(Z)P^{do}(\bar{X})P^{do}(X|\bar{X})P^{do}(\bar{Y}|X, Z) \right\} \tag{40}$$

Since the generative processes of all variables except for $\bar{X}$ and X are invariant before and after macro-intervention on $\bar{X}$, write

$$P(Z) = P(Z) = P^{do}(Z) \tag{41}$$

$$P(Y|X, Z) = P(Y|X, Z) = P^{do}(Y|X, Z) \tag{42}$$

Therefore, $P(\bar{Y}|X, Z) = P^{do}(\bar{Y}|X, Z)$. So write:

$$P(\bar{Y}|X, Z) = P(\bar{Y}|X, Z) = P^{do}(\bar{Y}|X, Z) \tag{43}$$

Note the decompositions:

$$P(Y|X) = \sum_j^{|\mathcal{Z}|} P(Y|X, Z = z_j)P(Z = z_j|X) \tag{44}$$

$$P^{do}(Y|X) = \sum_j^{|\mathcal{Z}|} P(Y|X, Z = z_j)P^{do}(Z = z_j) \tag{45}$$

$$P(\bar{Y}|X) = \sum_j^{|\mathcal{Z}|} P(\bar{Y}|X, Z = z_j)P(Z = z_j|X) \tag{46}$$

$$P^{do}(\bar{Y}|X) = \sum_j^{|\mathcal{Z}|} P(\bar{Y}|X, Z = z_j)P^{do}(Z = z_j) \tag{47}$$

$$P(\bar{Y}|\bar{X}) = \sum_{i,j}^{|\mathcal{X}|,|\mathcal{Z}|} P(\bar{Y}|X = x_i, Z = z_j)P(X = x_i, Z = z_j|\bar{X}) \tag{48}$$

$$P^{do}(\bar{Y}|\bar{X}) = \sum_{i,j}^{|\mathcal{X}|,|\mathcal{Z}|} P(\bar{Y}|X = x_i, Z = z_j)P^{do}(X = x_i, Z = z_j|\bar{X}) \tag{49}$$

$$\tag{50}$$

Define the following linear maps:

$$f_{1,x} : \mathbb{R}^{|\mathcal{Z}|} \longrightarrow \mathbb{R}^{|\mathcal{Y}|} \tag{51}$$

$$\mathbf{v} \longmapsto \sum_j^{|\mathcal{Z}|} P(Y|X=x, Z=z_j)v_j \tag{52}$$

$$f_{2,x} : \mathbb{R}^{|\mathcal{Z}|} \longrightarrow \mathbb{R}^{|\bar{\mathcal{Y}}|} \tag{53}$$

$$\mathbf{v} \longmapsto \sum_j^{|\mathcal{Z}|} P(\bar{Y}|X=x, Z=z_j)v_j \tag{54}$$

$$f_3 : \mathbb{R}^{|\mathcal{X}|\times|\mathcal{Z}|} \longrightarrow \mathbb{R}^{|\bar{\mathcal{Y}}|} \tag{55}$$

$$M \longmapsto \sum_{i,j}^{|\mathcal{X}|,|\mathcal{Z}|} P(\bar{Y}|X=x_i, Z=z_j)M_{ij} \tag{56}$$

Then conditions 1, 2, 3 are equivalent to

$$\exists x \in \mathcal{X}, \forall \bar{x} \in \bar{\mathcal{X}} : \tag{57}$$

$$P(Z|X=x) - P^{do}(Z) \notin Ker(f_{1,x}) \tag{58}$$

$$P(Z|X=x) - P^{do}(Z) \notin Ker(f_{2,x}) \tag{59}$$

$$P(X, Z|\bar{X}=\bar{x}) - P^{do}(X, Z|\bar{X}=\bar{x}) \in Ker(f_3) \tag{60}$$

Now analyse $P(X, Z|\bar{X})$ and $P^{do}(X, Z|\bar{X})$.

$$P(X, Z|\bar{X}) = \frac{P(\bar{X}|X)P(X|Z)P(Z)}{\sum_{i,j}^{|\mathcal{X}|,|\mathcal{Z}|} P(Z=z_j)P(X=x_i|Z=z_j)P(\bar{X}|X=x_i)} \tag{61}$$

$$P^{do}(X, Z|\bar{X}) = P(Z)P^{do}(X|\bar{X}) \tag{62}$$

An aside: A probability distribution of a discrete random variable, say $P(A), A \in \mathcal{A}, |\mathcal{A}| < \infty$ can be viewed as a finite-dimensional vector $\mathbf{v}$ such that $v_i = P(A = a_i)$. From now on we refer to this as the vector of $P(A)$.

Note that for a given $\bar{x}$, $P(\bar{X} = \bar{x}|X = x)$ and $P^{do}(X = x|\bar{X} = \bar{x})$ must be zero when $x \notin \pi_X^{-1}(\bar{x})$. Since $P(\bar{X}|X)$ is a factor of $P(X, Z|\bar{X})$, and $P^{do}(X|\bar{X})$ is a factor of $P^{do}(X, Z|\bar{X})$, we can work out the elements of the vectors $P(X, Z|\bar{X} = \bar{x})$ and $P^{do}(X, Z|\bar{X} = \bar{x})$ which are allowed to be non-zero i.e. precisely the elements corresponding to x with $x \in \pi_X^{-1}(\bar{x})$. Note that $\pi^{-1}(\bar{x}) \cap \pi^{-1}(\bar{x}') = \emptyset$ when $\bar{x} \neq \bar{x}'$.

Order the elements of $\bar{\mathcal{X}}$ as $\{\bar{x}_1, \cdots, \bar{x}_K, \ K = |\bar{\mathcal{X}}|\}$. Since $|\bar{\mathcal{X}}| < |\mathcal{X}|$, we can choose $\bar{x}$ such that $|\pi^{-1}(\bar{x})| > 1$. Wlog, let this be $\bar{x}_1$. Let $P(X^+, Z|\bar{X} = \bar{x}_1)$ and $P^{do}(X^{+,1}, Z|\bar{X} = \bar{x}_1)$ be matrices which contain the elements of $P(X, Z|\bar{X} = \bar{x}_1)$ and $P^{do}(X, Z|\bar{X} = \bar{x}_1)$ such that $X \in \pi^{-1}(\bar{x}_1)$. Precisely, $X^{+,1}$ is a vector of length $|\pi^{-1}(\bar{x}_1)|$ where $x_i^{+,1} \in \pi^{-1}(\bar{x}_1) \ \forall i = 1, \cdots, |\pi^{-1}(\bar{x}_1)|$. The ij^{th} element of $P(X^{+,1}, Z|\bar{X} = \bar{x}_1)$ is given by $P(X = x_i^{+,1}, Z = z_j|\bar{X} = \bar{x}_1)$. Choose $P(Z)$ and $P(X|Z)$, such that there exist some values of $x^{+,1}$ such that $P(X = x^{+,1}|\bar{X} = \bar{x}) \neq P^{do}(X = x^{+,1}|\bar{X} = \bar{x})$ and $P(Z|X = x^{+,1}) \neq P(Z)$. Wlog, let this be $x_1^{+,1}$. Then $P(Z|X = x_1^{+,1})$ and $P(Z)$ are linearly independent since they are both

normalised. It follows that $\mathbf{u}_{x_1^{+,1}} := P(Z|X^{+,1} = x_1^{+,1}) - P(Z)$ is linearly independent of $\mathbf{v}_{x_1^{+,1}} := P(Z|X = x_1^{+,1})P(X = x_1^{+,1}|\bar{X} = \bar{x}_1) - P(Z)P^{do}(X = x_1^{+,1}|\bar{x}_1)$. By Lemma 5, there exist (a distribution which we call) $P(\bar{Y}|X = x_1^{+,1}, Z)$ which maps $\mathbf{v}_{x_1^{+,1}}$ to 0 and maps $\mathbf{u}_{x_1^{+,1}}$'s to a non-zero vector. For the other values of $x_i^{+,k} \in \pi^{-1}(\bar{x}_k), \quad i, k \neq 1$, we only need to satisfy (60). Lemma 5 also immediately implies that there is $P(\bar{Y}|X = x_i^{+,k}, Z)$ such that $\mathbf{v}_{x_i^{+,k}} := P(Z|X = x_i^{+,k})P(X = x_i^{+,k}|\bar{X} = \bar{x}_k) - P(Z)P^{do}(X = x_i^{+,k}|\bar{X} = \bar{x}_k)$ is mapped to 0. Therefore, condition 3 and the second half of condition 1 are satisfied.

It remains to satisfy the first half of condition 1. This is easy, because the linear map $f_{2,x}$ is given by composing $f_{1,x}$ and $\mathbf{v} \in \mathbb{R}^{|\mathcal{Y}|} \mapsto \sum_k^{|\mathcal{Y}|} P(\bar{Y}|Y = y_k)v_k$. For a fixed $f_{2,x}$ and $P(\bar{Y}|Y)$, $\exists f_{1,x}$ s.t. $f_{2,x} = P(\bar{Y}|X) \circ f_{1,x}$, since, for example, take $P(Y = y_i|X, Z) = \frac{1}{|\pi^{-1}(\bar{y})|}P(\bar{Y} = \bar{y}|X, Z)$ for any $y_i \in \pi^{-1}(\bar{y})$. Then since f_{2,x^+} maps $\mathbf{u}_{x^+}$ to a non-zero vector, so must f_{1,x^+}. ∎

E.3. Section 4.1

Proof of Lemma 10. Implementing $do(\bar{y})$ according to $P(Y|\bar{y})$ results in the downstream effect $\sum_y P(\bar{Z}|y)P(y|\bar{y}) = \sum_y P(\bar{Z}|y, \bar{y})P(y|\bar{y}) = P(\bar{Z}|\bar{y})$. Micro-realization via $P(Y|\bar{y}, \bar{x})$ results in $\sum_y P(\bar{Z}|y)P(y|\bar{y}, \bar{x}) = \sum_y P(\bar{Z}|y, \bar{y}, \bar{x})P(y|\bar{y}, \bar{x}) = P(\bar{Z}|\bar{y}, \bar{x})$. The statement $P(\bar{Z}|\bar{y}) = P(\bar{Z}|\bar{y}, \bar{x})$ holds for all $\bar{x}, \bar{y}$ iff. (12) is true. ∎

E.4. Section 4.2

Proof of Lemma 11. For fixed $\bar{x}$, (14) states that $\bar{Z}$ is a function of noise that is independent of Y, thus we have equality of the interventional and observational probabilities $P(\bar{Z}|do(y), \bar{x}) = P(\bar{Z}|y, \bar{x})$. We conclude $P(\bar{Z}|do(\bar{y}), \bar{x}) = \sum_y P(\bar{Z}|do(y), \bar{x})p(y|\bar{y}, \bar{x}) = \sum_y P(\bar{Z}|y, \bar{x})p(y|\bar{y}, \bar{x}) = P(\bar{Z}|\bar{y}, \bar{x})$. ∎

Proof of Theorem 4.1. Immediate by replacing Y with X_i, Z with X_j, and X with B in the proof of Lemma 11 (note that the back-door condition for $\bar{G}$ implies the one for G because there the former can only differ by additional arrows. ∎

Appendix F. Auxiliary lemmas

Lemma 1 *For a given* $\boldsymbol{\alpha} \in \mathbb{R}^N$, $\boldsymbol{\alpha} \not\parallel (1, \cdots, 1)$ *and* $c \in \mathbb{R}$*, there exist an invertible* $N \times N$ *matrix* $H \in \mathbb{R}^{N \times N}$ *such that* $H_{1,i} = 1 \ \forall i = 1, \cdots, N$ *and* $\sum_i \alpha_i H_{i,1}^{-1} = c$*. Moreover, for any* $\boldsymbol{\Delta}$ *such that* $\sum_i \Delta_i = 1$*, there also exist* H *such that* $\boldsymbol{\Delta} = H_{:,1}^{-1}$*.*

Proof We aim to choose (column vectors) $\mathbf{u}_1, \cdots, \mathbf{u}_N$ and $\mathbf{v}_1, \cdots, \mathbf{v}_N$ such that $H = \begin{pmatrix} \mathbf{u}_1^\top \\ \vdots \\ \mathbf{u}_N \end{pmatrix}$ and $H^{-1} = (\mathbf{v}_1, \cdots, \mathbf{v}_N)$ satisfy the conditions in the claim. Let $\langle \cdot, \cdot \rangle$ denote the standard dot product in $\mathbb{R}^N$. First choose $\mathbf{u}_1 = (1, \cdots, 1)$.

Choose $\mathbf{v}_1^0 \in \mathbb{R}^N$ such that $\langle \mathbf{v}_1^0, \mathbf{u}_1 \rangle = 1$. Let $d = \langle \mathbf{v}_1^0, \boldsymbol{\alpha} \rangle$. If $d = c$, then set $\mathbf{v}_1 = \mathbf{v}_1^0$. Else, take $\mathbf{v} \in \mathbf{u}_1^\perp$ and $\mathbf{v} \notin \boldsymbol{\alpha}^\perp$. $\mathbf{v}$ exists since $\boldsymbol{\alpha} \not\parallel \mathbf{u}_1$. Choose $\mathbf{v}_1 = \mathbf{v}_1^0 + \frac{c-d}{\langle \mathbf{v}, \boldsymbol{\alpha} \rangle} \cdot \mathbf{v}$, then $\langle \mathbf{v}_1, \boldsymbol{\alpha} \rangle = \langle \mathbf{v}_1^0, \boldsymbol{\alpha} \rangle + \frac{c-d}{\langle \mathbf{v}, \boldsymbol{\alpha} \rangle} \cdot \langle \mathbf{v}, \boldsymbol{\alpha} \rangle = d + c - d = c$.

Now choose $\mathbf{u}_2, \cdots, \mathbf{u}_N$ s.t. i) $\langle \mathbf{u}_i, \mathbf{v}_1 \rangle = 0$, $\forall i = 2, \cdots, N$, and ii) $\mathbf{u}_2 \cdots, \mathbf{u}_N$ are linearly independent. This can be done since $\dim(\mathbf{v}_1^\perp) = N - 1$.

Note also that $\mathbf{u}_1, \cdots, \mathbf{u}_N$ are linearly independent: suppose $\sum_i m_i \mathbf{u}_i = \mathbf{0}$, then $\sum m_i \langle \mathbf{u}_i, \mathbf{v}_1 \rangle = 0$. But $\langle \mathbf{u}_i, \mathbf{v}_1 \rangle = 0$ for $i = 2, \cdots, N$; therefore, $m_1 \langle \mathbf{u}_1, \mathbf{v}_1 \rangle = 0$. Since $\langle \mathbf{u}_1, \mathbf{v}_1 \rangle = 1$, we deduce $m_1 = 0$. But this means $\sum_{i=2}^N m_i \mathbf{u}_i = \mathbf{0}$. Since we know $\mathbf{u}_2, \cdots, \mathbf{u}_N$ are linearly independent, $m_i = 0$ for $i = 2, \cdots, N$.

For $j = 2, \cdots, N$, choose $0 \neq \tilde{\mathbf{v}}_j \in \mathrm{Span}\left(\{\mathbf{u}_i, i \neq j\}\right)^\perp$. Since $\{\mathbf{u}_i, i \neq j\}$ are linearly independent, $\dim(\mathrm{Span}\left(\{\mathbf{u}_i, i \neq j\}\right)^\perp) = 1$, and thus $\mathrm{Span}(\tilde{\mathbf{v}}_j) = \mathrm{Span}\left(\{\mathbf{u}_i, i \neq j\}\right)^\perp$. Then $\langle \tilde{\mathbf{v}}_j, \mathbf{u}_i \rangle = 0$ when $i \neq j$. When $i = j$, $\langle \tilde{\mathbf{v}}_j, \mathbf{u}_j \rangle \neq 0$ since otherwise $\mathbf{u}_j \in \tilde{\mathbf{v}}_j^\perp = \mathrm{Span}\left(\{\mathbf{u}_i, i \neq j\}\right)$. Therefore, choose $\mathbf{v}_j = \frac{\tilde{\mathbf{v}}_j}{\langle \tilde{\mathbf{v}}_j, \mathbf{u}_j \rangle}$.

The above shows the first statement.

The second statement is obvious by observing that $\mathbf{v}_1$ can be chosen as $\mathbf{\Delta}$ by construction. ∎

Lemma 5 *Let $\mathbf{v} \in \mathbb{R}^{M \times N}$ be a vector with $\sum_{i,j} v_{ij} = 0$, and $\mathbf{u}_i \in \mathbb{R}^N$ a vector such that $\sum_j u_{i,j} = 0$, where $u_{i,j}$ is the jth element of $\mathbf{u}_i$. Suppose $\mathbf{u}_i$ and $\mathbf{v}_i$ are linearly independent, $M \geq 3$, $N \geq 2$, $K \geq 1$. Then there exists a distribution of discrete variables $P(\bar{Y}|X, Z)$ where $|\mathcal{X}| = M$, $|\mathcal{Z}| = N$, $|\bar{\mathcal{Y}}| = K$ such that*

1. $\sum_{i,j=1}^{M,N} P(\bar{Y}|X = x_i, Z = z_j) v_{ij} = \mathbf{0}$.

2. $\sum_j^N P(\bar{Y}|X = x_i, Z = z_j) u_{i,j} \neq \mathbf{0}$.

Proof

Let P_{ij}, $i = 1, \cdots, M$, $j = 1, \cdots, N$ denote the vector of $P(\bar{Y}|X = x_i, Z = z_j)$.

The case with $v_{ij} = 0$ for all i, j is trivial, so assume that there exist some (i, j) such that $v_{ij} \neq 0$.

If there is some i, j such that $v_{ij} = 0$, then we are done by choosing $P_{i_1, j_1} = P_{i_2, j_2}$ for any $(i_1, j_1), (i_2, j_2) \neq (i, j)$ as any valid probability vector, and choosing P_{ij} to be a different valid probability vector with all elements strictly between 0 and 1. For example, we can choose $P_{i_1, j_1} = (1/K, \cdots, 1/K)$, and $P_{ij} = (1/K + \epsilon, 1/K - \epsilon, 1/K \cdots, 1/K)$ with a small enough ϵ.

If $v_{ij} \neq 0$ for all (i, j), then clearly $\sum_{(i,j) \in \mathcal{I}_+} v_{ij} = -\sum_{(i,j) \in \mathcal{I}_-} v_{ij}$, where $\mathcal{I}_+$ is the index set for the positive v_{ij}'s and $\mathcal{I}_-$ are the negatives. For ease of notation let $S := \sum_{(i,j) \in \mathcal{I}_+} v_{ij} = -\sum_{(i,j) \in \mathcal{I}_-} v_{ij}$.

Now first assign arbitrary probability vectors for each P_{ij}, $(i, j) \in \mathcal{I}_+$. Wlog, we can assign different probability vectors to each, ensuring that each element of the vectors are strictly between 0 and 1. Then the linear combination $\sum_{(i,j) \in \mathcal{I}_+, k} v_{ij} P_{k,ij}$ is equal to S; here, $P_{k,ij}$ is the kth element of the vector of $P_{k,ij}$. Now assign for each $(i, j) \in \mathcal{I}_-$, $P_{ij} := \frac{1}{S} \sum_{(i,j) \in \mathcal{I}_+} v_{ij} P_{ij}$. Clearly, every element of P_{ij} is non-negative and all elements add up to 1, so P_{ij} is a probability vector. Moreover, $P(\bar{Y}|X, Z)$ constructed like this will satisfy the requirement that $\sum_{i,j=1}^{M,N} P(\bar{Y}|X = x_i, Z = z_j) v_{ij} = 0$.

Now it remains to check condition 2. Take a fixed $i \in \{1, \cdots, M\}$, if $\sum_{i,j} u_{ij} P_{ij} = \mathbf{0}$, then consider $\mathbf{v}_i$ and $\mathbf{u}_i$. Since they are linearly independent vectors, the linear map induced by the usual inner product with $\mathbf{v}_i$ has a kernel with a subspace of at least dimension one that is not contained in the kernel of the linear map induced by $\mathbf{u}_i$. Let $\mathbf{d}$ be a unit vector in this subspace. Define

$P_i = [P_{i1}, \cdots, P_{iN}]$. P_i is a $K \times N$ matrix. Construct P'_i by subtracting a small enough multiple of **d** from the first row of P_i and adding the same multiple to the second row of P_i. Redefine $P_i := P'_i$, and we will have $\sum_j u_{ij} P_{ij} \neq \mathbf{0}$ and $\sum_{i,j} v_{ij} P_{ij} = \mathbf{0}$. Repeat this for every $i = 1, \cdots, M$ and conditions 1 and 2 will both be satisfied.

∎

Appendix G. Examples

G.1. Hours of work and pay

For manual labourers in a manufacturing plant, the number of hours they work (X_i) determines their monthly wage (Y_i). In particular, the monthly wage (Y_i) of a manual labourer is a product of the number of hours they work (X_i) and their hourly wage (α_i):

$$Y_i = \alpha_i X_i \tag{63}$$

The hourly wage may differ between labourers due to various factors like work experience. To get a high-level overview of the human resources in the manufacturing plant, its business owner keeps track of total work hours ($\bar{X}$) and total cost of wages ($\bar{Y}$) on a monthly basis. That is, we have a summation coarsening map:

$$\bar{X} = \sum X_i, \qquad \bar{Y} = \sum Y_i$$

In this scenario, Gaussian distribution reasonably approximates the distribution of the number of hours of work of a manual labourer (X_i) as most labourers work around an average number of hours, with few working extra hours more or less. Moreover, it is reasonable to assume that the number of hours of work of each labourer is independent of other labourers. Altogether, we have the linear unconfounded case with independent Gaussian micro-level causes.

G.2. Genomic micro-array data for prediction of diseases

Genomic micro-array data contains personal genome information. A micro-array is a rectangular grid where every column contains the genomic expressions of one person (or subject), and every row contains the expressions of one coding gene for every person. Personal genomic expressions as collected in the micro-array are used to predict certain disease types (Y). The number of genes (X_i) present on a micro-array is vast; moreover, a population of genes could be jointly causing a disease. This may necessitate coarsening in drug design.

While two drugs may both claim to intervene on the causes of disease Y on a population level, $\pi((X_i)_i)$, it is possible that they impact the individual amounts of each gene expression differently. Moreover, due to the complexity of gene regulatory networks, there could be confounding present at the micro level. Suppose the practitioners are willing to assume a Linear Gaussian model, then our example in 3.2.1 suggests that there is a way for the drug to intervene on the macro variable such that the causal influence of the drug can be directly read off from observational data.

G.3. Political campaigning for votes

During the running up to the general election, politicians go around the country to campaign for votes. The amount of time they spend campaigning at county i can be written X_i. Some (simplistic) political model may assume a linear relationship between the amount of time campaigning in county i and the votes harvested at that county Y_i:

$$Y_i = \alpha_i X_i \tag{64}$$

At the end of the campaign, the team may calculate how much effort they put in, in terms of time, and how many votes they won in total.

$$\bar{X} = \sum X_i, \quad \bar{Y} = \sum Y_i \tag{65}$$

It will be noticed that different counties responds differently to the campaign, so α_i is not constant. Now, different campaign teams may all decide to increase their campaign time in order to get more votes, but the resultant vote changes may still be different due to the different allocations to each county.

Appendix H. More Detailed Discussion of Related Work

(Rubenstein et al., 2017) propose exact transformations, following which (Beckers and Halpern, 2019) propose a series of stronger notions of causal abstraction. As illustrated by the example in Example 1, the condition required to call a macro system an exact transformation of the the fine-grained one is not satisfied even in extremely simple cases. The moral reason being that when micro causal variables are aggregated, the causal relationships between the micro causal variables do not in general align with the aggregation, such that if multiple micro-cause-states get coarsened to the same macro-state, then the corresponding micro-effect-states get coarsened also to the same macro-state. By contrast, this work explores what happens when the macro system cannot be called an exact-transformation of the micro system.

(Beckers et al., 2019) explores how to quantify the approximation error when the abstraction is not exact. We take a different angle in this work and do not deal with how good our macro-model is at approximating the micro-model. Rather, our work observes that the aggregated system loses resolution on the original, micro-system, and analyses properties when different pairings between the macro- and micro-systems are made.

A different but closely-related line of work is undertaken by (Chalupka et al., 2016; Rischel and Weichwald, 2021), where the coarsest possible consistent coarsening is sought. The macro-variable states are constructed as equivalence classes (of micro-variable states) in the paper, and every micro-variable state in the same equivalence class leads to the same micro-variable effect - this ensures that interventions on the equivalence classes are well-defined. Our work lies in the realm beyond the coarsest consistent partition, and in these cases, the macro interventions become ill-defined.

Beckers et al. (2023) provides a framework to include constraints and structural equation model in a unified model. Macro variables in their framework leads to constraints on the system, and they propose to think of interventions on macro variables as implemented by disconnecting some of the other variables in the constraint from their original structural equations. Different ways of implementations can arise from choosing different variables to disconnect. This is similar to our notion of micro-realisations, and the differences are: 1) Beckers et al. (2023) do not consider soft

micro-level interventions as micro-realisations whereas we do, 2) they do not define causal models over only the macro variables whereas we do; in doing so, we explicate consequences of aggregation such as ambiguity of confounding.

Anand et al. (2023) describes a coarse-graining of only the causal DAG, and generalises do-calculus to deal with clusters of variables. However, they do not aggregate the variables within the cluster. As a result, ambiguities do not arise from different micro-realisations in the corresponding micro-variables of a particular macro-intervention.

Rischel and Weichwald (2021) proposes a framework of deriving approximate causal abstractions such that the errors are composable. However, their perspective differs from ours as they consider a probability distribution of different micro realisations of a given macro intervention, and they use this to define their approximation error for a given macro model, whereas we define macro models based on micro-realisations.

Zennaro et al. (2023) introduces the pseudo inverse of the function mapping from the micro-variable states and macro-variable states. This allows them to generalise measures of approximation error to be along other paths on the commutative diagram. The pseudo inverse will, given a macro-variable states as input, output a vector of equal weights one for each micro state that gives rise to the macro state. In the case of a uniform distribution over the micro states, their pseudo inverse a special case of our natural micro realisation. Then, they learn their macro relation based on the given pseudo inverse, whereas we consider different macro relations which are entailed by different micro implementations; thus our perspective is complementary to theirs.

Finally, our work inspires assosiation with (Blom et al., 2019), which describes a framework for modelling cyclic causal models, where solutions to the model are the equilibrium states for some initial conditions, and interventions amount to changing these initial conditions. In particular, the causal dynamics of the model are fixed by differential equations. Although some might be tempted to view the relationship between micro and macro variables in our case as cyclic, it is, however, fundamentally different. In our case by specifying the $P^{do}_{X|\bar{x}}$ in the macro-intervention, the dynamics between the macro and micro variables are freely customisable by the practitioner, whereas in their case, the dynamics between the macro and micro variables are fixed a priori by some constraints, for example, by the laws of physics. That being said, there could be situations where the dynamics fixed by the a priori constraints, in some canonical way, already gives rise to the confounding-inhibiting interventions that we described.