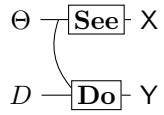
# Causal Statistical Decision Theory|What are interventions?

David Johnston

August 12, 2020

Note on terminology: I am trying the name "See-Do model" to describe the folowing:




I was calling it a "causal theory" before. Reasons for the change: I think "See-Do" helps to understand what the model does, and the name doesn't make premature claims to explain causality. Also, it's only two syllables which I like.

## 1 Why you need decisions

CSDT differs from Causal Bayesian Networks in that See-Do models require a set $D$ of decisions to be specified as part of the definition, just as a function requires a domain to be specified as part of its definition. Causal Bayesian Networks, in contrast, appear to define causal effects of variables without reference to an underlying set of decisions. We show that this distinction is only apparent, and that Causal Bayesian Networks also require the specification of "causal atoms" which correspond precisely to the decision set in a See-Do model.

In what follows, I will call the approach of CSDT a "decisions first" approach, as a set $D$ of decisions must be chosen before anything can be said about causation. I will call the approach taken by Causal Bayesian Networks a "causes first" approach, as one could take the view that the investigation of consequences of decisions is a specialisation of the study of causal effect, and such is my understanding of the philosophy behind the Causal Bayesian Network approach.

"Causes first" and "decisions first" approaches do not clearly necessitate one follow a particular school of *modelling* causal relationships. *Influence diagrams* are CBN-like graphical models that follow a decisions first paradigm (Peters et al., 2017; Woodward, 2016; Dawid, 2002). It is unclear where the Potential Outcomes school sits on this question - like CBNs, causal effects in the Potential Outcomes school talks about causal effects of variables, an approach that avoids explicit definition of a set of decisions. On the other hand, unlike CBN models,

Potential Outcomes models typically only consider a subset of variables to have causal effects, and consider causal effects to be difficult to define in general (Rubin, 2005), so *some* choices are made in terms of which variables can underwrite causal effects.

An apparent difference between causes first and decisions first paradigms is whether the question "what is the causal effect of $\mathsf{X}$ on $\mathsf{Y}$?" is, in general, well-defined. In the decisions first approach, it seems to be necessary to ask "what is the causal effect of $\mathsf{X}$ on $\mathsf{Y}$ with respect to decisions $D$?", and the question is ill-posed without this clarification. On the other hand, it seems to be well defined in the causes first approach.

However, this too is merely an *apparent* difference. Because the causes first approach requires the specification of a set of causal atoms, to properly pose the question in the causes first paradigm requires the clarification "what is the causal effect of $\mathsf{X}$ on $\mathsf{Y}$ *with respect to atoms* $A$?".

I propose that the actual difference here is the hypothesis of *causal universality*: that there exists a unique set of causal atoms $A^*$ that is appropriate for every problem. If such a set exists then the original question "what is the causal effect of $\mathsf{X}$ on $\mathsf{Y}$?" can be understood as implicitly invoking $A^*$. An analogous hypothesis of *decision universality* can do the same thing in the decisions first paradigm.

I think there might be a useful theorem that work well with universality, but not sure yet

And I discuss whether or not you might want to assume universality

## 1.1 Structural Causal Models

We adopt the framework of cyclic structural causal models to make our arguments, adapted from Bongers et al. (2016). This is somewhat non-standard, but allows us to make a stronger argument for the impossibility of modelling arbitrary sets of variables using structural interventional models.

**Definition 1.1** (Structural Causal Model)**.** A structural causal model (SCM) is a tuple

$$\mathcal{M} := \langle \mathcal{I}, \mathcal{J}, \mathbf{X}_{\mathcal{I}}, \mathbf{E}_{\mathcal{J}}, \mathbf{f}_{\mathcal{I}}, \mathbb{P}_{\mathcal{E}}, \mathbf{E}_{\mathcal{J}} \rangle \tag{1}$$

where

- $\mathcal{I}$ is a finite index set of *endogenous variables*
- $\mathcal{J}$ is a finite index set of *exogenous variables*
- $\mathbf{X}_{\mathcal{I}} := \{X_i\}_{\mathcal{I}}$ where, for each $i \in \mathcal{I}$, $(X_i, \mathcal{X}_i)$ is a standard measurable space taking and the codomain of the $i$-th endogenous variable
- $\mathbf{E}_{\mathcal{J}} := \{E_j\}_{\mathcal{J}}$ where, for $j \in \mathcal{J}$, $E_j$ is a standard measurable space and the codomain of the $j$-th endogenous variable

- $\mathbf{f}_{\mathcal{I}} = \underline{\otimes}_{i \in \mathcal{I}} f_i$ is a measurable function, and $f_i : \mathbf{X}_{\mathcal{I}} \times \mathbf{E}_{\mathcal{J}} \to X_i$ is the causal mechanism controlling $\mathsf{X}_i$
- $\mathbb{P}_{\mathcal{E}} \in \Delta(\mathbf{E}_{\mathcal{J}})$ is a probability measure on the space of exogenous variables
- $\mathbf{E}_{\mathcal{J}} = \underline{\otimes}_{j \in \mathcal{J}} \mathsf{E}_j$ is the set of exogenous variables, with $\mathbb{P}_{\mathcal{E}} = \mathbf{E}_{\mathcal{J}\#} P_{\mathcal{E}}$ and $\mathsf{E}_j$ is the j-th exogenous variable with marginal distribution given by $\mathsf{E}_{j\#}\mathbb{P}_{\mathcal{E}}$

If for $\mathbb{P}_{\mathcal{E}}$-almost every $\mathbf{e} \in \mathbf{E}_{\mathcal{J}}$ there exists $\mathbf{x} \in \mathbf{X}_{\mathcal{I}}$ such that

$$\mathbf{x} = \mathbf{f}_{\mathcal{I}}(\mathbf{x}, \mathbf{e}) \tag{2}$$

Then an SCM $\mathcal{M}$ induces a unique probability space $(\mathbf{X}_{\mathcal{I}} \times \mathbf{E}_{\mathcal{J}}, \mathcal{X}_{\mathcal{I}} \otimes \mathcal{E}_{\mathcal{J}}, \mathbb{P}_{\mathcal{X} \otimes \mathcal{E}})$ (Bongers et al., 2016). If no such solution exists then we will say an SCM is invalid, as it imposes mutually incompatible constraints on the endogenous variables.

If an SCM induces a unique probability space then there exist random variables $\{\mathsf{X}_i\}_{i \in \mathcal{I}}$ such that, almost surely Bongers et al. (2016):

$$\mathsf{X}_i = f_i(\mathbf{X}_{\mathcal{I}}, \mathbf{E}_{\mathcal{J}}) \tag{3}$$

Where $\mathbf{X}_{\mathcal{I}} = \underline{\otimes}_{i \in \mathcal{I}} \mathsf{X}_i$.

A structural causal model can be transformed by *mechanism surgery*. This is an operation that replaces a model $\mathcal{M}$ with a modified model $\mathcal{M}'$ that shares all elements except the set of causal mechanisms, which is replaced by an arbitrarily different set $\mathbf{f}'_{\mathcal{I}}$.

If the mechanism surgery that transforms $\mathcal{M}$ to $\mathcal{M}'$ is such that $f'_i \neq f_i$ and $f'_j = f_j$, $j \neq i$, then we say that $\mathcal{M}'$ represents *an intervention on* $\mathsf{X}_i$. If $f'_i$ is furthermore a constant function equal to $x$ then we say that $\mathcal{M}'$ represents a *hard intervention on* $i$ and use the special notation $\mathcal{M}^{do(\mathsf{X}_i = x)} := \mathcal{M}'$. Similarly, $\mathcal{M}^{do(\mathsf{X}_i = x, \mathsf{X}_j = y)}$ is a model representing a hard intervention on $i$ and $j$. The *causal effect* of $\mathsf{X}_i$ is the set of hard interventional SCMs $\{\mathcal{M}^{do(\mathsf{X}_i = x)} | x \in X_i\}$, which is generated by the fundamental model $\mathcal{M}$ and the operation of hard intervention on $\mathsf{X}_i$.

We say a *causal model* is any kind of model that defines causal effects. An SCM $\mathcal{M}$ in combination with hard interventions defines causal effects, so an SCM is a causal model. Call each interventional model $\mathcal{M}^{do(\mathsf{X}_i = x)}$ a *submodel* of $\mathcal{M}$.

Strictly, the random variables $\mathsf{X}_i$ depend on the probability space induced by a particular model $\mathcal{M}$, they are intended to refer to "the same variable" across different models that are related by mechanism surgery. We will abuse notation and use $\mathsf{X}_i$ to refer to the *family* of random variables induced by a set of models related by mechanism surgery, and rely on explicitly noting the measure $\mathbb{P}_{...}(...)$ to specify exactly which random variables we are talking about.

In practice, we typically specify a "small" SCM containing a few endogenous variables $\mathcal{I}$ (called a "marginal SCM" by Bongers et al. (2016)) which is understood to summarise the relevant characteristics of a "large" SCM containing many variables $\mathcal{I}^*$. We will argue that without restrictions on the large set of variables $\mathcal{I}^*$, surgically transformed SCMs will usually be invalid.

Incidentally, this messiness with random variables can be solved if we use See-Do models.

## 1.2 Not all variables have well-defined interventions

A long-running controversy about causal inference concerns the question of "the causal effect of body mass index on mortality". On the one hand, Hernán and Taubman (2008) and others claim that there is no well-defined causal effect of a person's body mass index (BMI), defined as their weight divided by their height, and their risk of death. Pearl claims, in defence of Causal Bayesian Networks, that the causal effect of *obesity* is well-defined, though it is not clear whether he defends the proposition that BMI itself has a causal effect:

> That BMI is merely a coarse proxy of obesity is well taken; obesity should ideally be described by a vector of many factors, some are easy to measure and others are not. But accessibility to measurement has no bearing on whether the effect of that vector of factors on morbidity is "well defined" or whether the condition of consistency is violated when we fail to specify the interventions used to regulate those factors. (Pearl, 2018)

We argue that BMI does *not* have a well-defined causal effect, and without further assumptions neither does any variable.

## 1.3 Necessary relationships

The relationship between a person's body mass index, their weight and their height defines what body mass index is. A fundamental claim of ours is that any causal model that defines "the causal effect of body mass index" should do so without reference to any submodel that violates this definitional relationship violation of the definition. This is an important assumption, and it rests on a judgement of what causal models ought to do. I think it is quite clear that when anyone asks for a causal effect, they expect that any operations required to define the causal effect *do not change the definitions of the variables they are employing.* While theories like SCMs have a role in sharpening our understanding of the term *causal effect*, the thing called a "causal effect" in an SCM should still respect some of our pre-theoretic intuitions about what causal effects are or else it should be called something else. "Causal effects" that depend on redefining variables do not respect pre-theoretic intuitions about what causal effects are:

- If I ask for the "causal effect of a person's BMI", I do not imagine that I am asking what would happen if someone's BMI were defined to be something other than their weight divided by their height
- If I ask for the "causal effect of a person's weight", I do not imagine that I am asking what would happen if someone's weight were not equal to their volume multiplied by their density
- If I ask for the "causal effect of a person's weight", I also do not imagine that I am asking what would happen if their weight were not equal to the weight of fat in their body plus the weight of all non-fat parts of their body

- If I ask for the "causal effect of taking a medicine", I do not imagine that I am asking what would happen if a person were declared to have taken a medicine independently of whatever substances have actually entered their body and how they entered

We will call relationships that have to hold *necessary relationships*. We provide the example of relationships that have to hold by definition as examples of relationships that should be necessary in causal models, but one might also wish to stipulate that certain laws of physics are required to hold in all submodels.

If an SCM contains variables that are necessarily related, we wish to impose the additional restriction that these necessary relationships hold for every submodel. This can be done by extending the previous definition:

**Definition 1.2** (SCM with necessary relationships)**.** An SCM with necessary relationships (SCNM) is a tuple $\mathcal{M} := \langle \mathcal{I}, \mathcal{J}, \mathbf{X}_{\mathcal{I}}, \mathbf{E}_{\mathcal{J}}, \mathbf{f}_{\mathcal{I}}, \mathbf{g}_{\mathcal{I}}, \mathbb{P}_{\mathcal{E}}, \mathbf{E}_{\mathcal{J}} \rangle$, which is an SCM with the addition of a vector function of *necessary relationships* $\mathbf{g}_{\mathcal{I}} := \underline{\otimes}_{i \in \mathcal{I}} g_i$ where each $g_i : \mathbf{X}_{\mathcal{I}} \to X_i$ is a necessary relationship involving $\mathsf{X}_i$.

An SCM with necessary induces a unique probability space if for $\mathbb{P}_{\mathcal{E}}$-almost every $e \in \mathcal{E}$ there exists a unique $\mathbf{x} \in \mathbf{X}_{\mathcal{I}}$ such that simultaneously

$$\mathbf{x} = \mathbf{f}_{\mathcal{I}}(\mathbf{x}, \mathbf{e}) \tag{4}$$
$$\mathbf{x} = \mathbf{g}_{\mathcal{I}}(\mathbf{x}) \tag{5}$$

If no such $\mathbf{x}$ exists then an SCNM is invalid.

Mechanism surgery for SCNMs involves modification of $\mathbf{f}_{\mathcal{I}}$ only, just like SCMs.

If we wish to stipulate that a particular variable $\mathsf{X}_i$ has no causal relationships or necessary relationships we can specify this with the trivial mechanisms $f_i : (\mathbf{x}, \mathbf{e}) \mapsto x_i$ and $g_i : \mathbf{x} \mapsto x_i$ respectively. An SCNM $\mathcal{M}$ with the trivial necessary relationship $\mathbf{g}_{\mathcal{I}} : \mathbf{x} \mapsto \mathbf{x}$ induces the equivalent probability spaces as the SCM obtained by removing $\mathbf{g}_{\mathcal{I}}$ from $\mathcal{M}$.

Because BMI is always equal height/weight, a causal model that includes height and weight ought to be able to model anything that a model containing height, weight and BMI can model.

explain why

However, as Theroem 1.4 shows, if an SCNM with height, weight and BMI can be derived from an SCNM containing just height and weight then there are no valid interventions on BMI.

**Definition 1.3** (Derived model)**.** Given a SCNM $\mathcal{M} := \langle \mathcal{I}, \mathcal{J}, \mathbf{X}_{\mathcal{I}}, \mathbf{E}_{\mathcal{J}}, \mathbf{f}_{\mathcal{I}}, \mathbf{g}_{\mathcal{I}}, \mathbb{P}_{\mathcal{E}}, \mathbf{E}_{\mathcal{J}} \rangle$, say $\mathcal{M}' = \langle \mathcal{I}', \mathcal{J}, \mathbf{X}_{\mathcal{I}'}, \mathbf{E}_{\mathcal{J}}, \mathbf{f}_{\mathcal{I}'}, \mathbf{g}_{\mathcal{I}'}, \mathbb{P}_{\mathcal{E}}, \mathbf{E}_{\mathcal{J}} \rangle$ is *derived* from $\mathcal{M}$ if there exists some additional index/variable/relationships $i' \notin \mathcal{I}, X_{i'}, f_{i'}, g_{i'}$ such that

$$\mathcal{I}' = \mathcal{I} \cup \{i'\} \tag{6}$$
$$\mathbf{X}_{\mathcal{I}'} = \mathbf{X}_{\mathcal{I}} \cup X_{i'} \tag{7}$$
$$\mathbf{f}_{\mathcal{I}'} = \mathbf{f}_{\mathcal{I}} \underline{\otimes} f_{i'} \tag{8}$$
$$\mathbf{g}_{\mathcal{I}'} = \mathbf{g}_{\mathcal{I}} \underline{\otimes} g_{i'} \tag{9}$$

Call $i', X_{i'}, f_{i'}, g_{i'}$ the *additional elements*.

I need to stipulate that $f_k$ "forgets" $\mathsf{X}_{i'}$ for $k \neq i'$, $g_k$ "forgets" for all $k$

**Theorem 1.4** (Interventions and necessary relationships don't mix)**.** *If $\mathcal{M}'$ is derived from $\mathcal{M}$ with the additional elements $i', X_{i'}, f_{i'}, g_{i'}$ and both $\mathcal{M}$ and $\mathcal{M}'$ are uniquely solvable and $\mathbb{P}_{\mathcal{X}'\otimes\mathcal{E}}(\mathsf{X}_{i'})$ is not single valued then no hard interventions on $\mathsf{X}_{i'}$ are possible.*

*Proof.* Because $\mathcal{M}$ is uniquely solvable, for $\mathbb{P}_{\mathcal{E}}$ almost every $\mathbf{e}$ there is a unique $\mathbf{x}^e$ such that

$$\mathbf{x}^e = \mathbf{f}_{\mathcal{I}}(\mathbf{x}^e, \mathbf{e}) \tag{10}$$
$$\mathbf{x}^e = \mathbf{g}_{\mathcal{I}}(\mathbf{x}^e) \tag{11}$$

Because $\mathcal{M}'$ is also uniquely solvable, for $\mathbb{P}_{\mathcal{E}}$ almost every $\mathbf{e}$ we have

$$x^e_{i'} = \mathbf{g}_{i'}(\mathbf{x}^e) \tag{12}$$

Because $\mathbb{P}_{\mathcal{X}'\otimes\mathcal{E}}(\mathsf{X}_{i'})$ is not single valued there are non-null sets $A, B \in \mathcal{E}$ such that $e_a \in A$, $e_b \in B$ implies

$$\mathbf{g}_{i'}(\mathbf{x}^{e_a}) \neq \mathbf{g}_{i'}(\mathbf{x}^{e_b}) \tag{13}$$

Therefore there exists no $a \in X_{i'}$ that can simultaneously satisfy 12 for almost every $\mathbf{e}$. However, any hard intervention $\mathcal{M}'^{,do(\mathsf{X}_{i'}=a)}$ requires such an $a$ in order to be solvable. □

**Corollary 1.5.** *Either there are no hard interventions defined on BMI or there is no SCNM containing height and weight with a unique solution from which an SCNM containing height, weight and BMI can be derived.*

I can formalise the following, but I'm just writing it out so I can get to the end for now

The problem posed by Theorem 1.4 can be circumvented to some extent by joint interventions. Consider the variables $\mathsf{X}_1$ and $\mathsf{X}_2$ where $\mathsf{X}_1 = -\mathsf{X}_2$ necessarily. While Theorem 1.4 disallows interventions on $\mathsf{X}_2$ individually (supposing we can obtain a unique model featuring only $\mathsf{X}_1$), it does not disallow interventions that jointly set $\mathsf{X}_1$ and $\mathsf{X}_2$ to permissible values. In this case, this is unproblematic as the only joint intervention that sets $\mathsf{X}_1$ to 1 must also set $\mathsf{X}_2$ to $-1$.

If we have non-invertible necessary relationships such as $\mathsf{X}_1 = \mathsf{X}_2 + \mathsf{X}_3$, however, there are now *multiple* joint interventions on $\mathsf{X}_1$ that can be performed. I regard this as the most plausible solution to the difficulties raised so far: for variables that are in non-invertible necessary relationships, there is a set of operations associated with the "intervention" that sets $\mathsf{X}_1 = 1$.

However, we still need to make sure the interventions that we have supposed comprise the operations associated with setting $\mathsf{X}_1 = 1$ exist themselves. It is sufficient that the SCNM with $\mathsf{X}_1$ is derived from a higher order *uniquely solvable SCM* with $\mathsf{X}_2$ and $\mathsf{X}_3$ only .

because interventions are defined in uniquely solvable SCMs and derivation preserves interventions on the old variables

And necessary? There might be "degenerate" necessary relationships that don't harm the possibility of defining interventions, and I'd need to show an equivalence to an SCM in this case

If any variables are included in a causal model that are necessarily related to other variables (and honestly, is there any variable that isn't?), it is not enough to suppose that the model being used is a marginalisation of some larger causal model. Rather, it must be obtained by derivation and marginalisation from some model that represents the basic interventions that are possible, which we call the *atomic model*.

**Definition 1.6** (Atomic model)**.** Given an SCNM $\mathcal{M}$, the *atomic model* $\mathcal{M}_{\text{atom}}$ is a uniquely solvable SCM such that there exists a model $\mathcal{M}$ is derived from of $\mathcal{M}_{\text{atom}}$.

Typically, in order to get an actually usable model you'll also need to marginalize, but I think this complication can be avoided

**Definition 1.7** (Causal universality hypothesis)**.** There exists a uniquely solvable SCM $\mathcal{M}_{\text{atom}}$ which is the atomic model that correctly represents all decision problems

what does that mean?

or causal problems?

I don't know how to define "correctly represents" or "causal problem", but it seems like something like the universality hypothesis is necessary if you want to define "the causal effect of X" independent of any atomic model

Relate decisions to interventions on atomic model. Decisions->atomic model is straightforward, but the reverse direction is not so obvious

Causal effects are uniquely defined via atoms iff they are defined via decisions

Are there any plausible ways to construct atomic models?

## References


Stephan Bongers, Jonas Peters, Bernhard Schölkopf, and Joris M. Mooij. Theoretical Aspects of Cyclic Structural Causal Models. *arXiv:1611.06221 [cs, stat]*, November 2016. URL `http://arxiv.org/abs/1611.06221`. arXiv: 1611.06221.

A. P. Dawid. Influence Diagrams for Causal Modelling and Inference. *International Statistical Review*, 70(2):161–189, 2002. ISSN 1751-5823. doi: 10.1111/j.1751-5823.2002.tb00354.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-5823.2002.tb00354.x`.

M. A. Hernán and S. L. Taubman. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *International Journal of Obesity*, 32(S3):S8–S14, August 2008. ISSN 1476-5497. doi: 10.1038/ijo.2008.82. URL `https://www.nature.com/articles/ijo200882`.

Judea Pearl. Does Obesity Shorten Life? Or is it the Soda? On Non-manipulable Causes. *Journal of Causal Inference*, 6(2), 2018. doi: 10.1515/jci-2018-2001. URL https://www.degruyter.com/view/j/jci.2018.6.issue-2/jci-2018-2001/jci-2018-2001.xml.

Jonas Peters, Dominik Janzing, and Bernard Schölkopf. *Elements of Causal Inference*. MIT Press, 2017.

Donald B. Rubin. Causal Inference Using Potential Outcomes. *Journal of the American Statistical Association*, 100(469):322–331, March 2005. ISSN 0162-1459. doi: 10.1198/016214504000001880. URL https://doi.org/10.1198/016214504000001880.

James Woodward. Causation and Manipulability. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016. URL https://plato.stanford.edu/archives/win2016/entries/causation-mani/.

## Appendix: