Title: Classical causal models in string diagrams

Speakers: Robin Lorenz

Series: Quantum Foundations, Quantum Information

Date: September 17, 2024 - 2:10 PM

URL: https://pirsa.org/24090116

Classical causal models in string diagrams

Robin Lorenz Sean Tull

QUANTINUUM

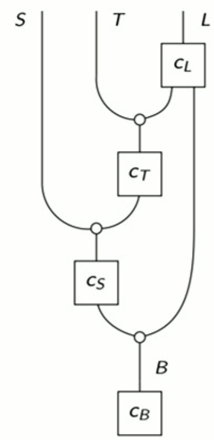
Causalworlds, Perimeter Institute, Sept 2024

What and why

Classical causal models

Pearl
Spirtes, Glymour, Scheines
⋮

... in string diagrams



Coecke, Spekkens (2012)
Jacobs, Zanasi (2016)
Cho, Jacobs (2016/2019)
Fritz (2020)
Fritz, Klingler (2022)

Fong (2013)
Jacobs, Kissinger, Zanasi (2019/2021)
Friend, Kissinger (2023)

Generalisation, formalisation (category theory)

Foundations classical causal inference

Differences quantum causal models

Intuitive, easy representation

(X)AI

Sean Tull

Causal models in string diagrams

Robin Lorenz*, Sean Tull†

Quantinuum, 17 Beaumont Street, Oxford, UK

Abstract

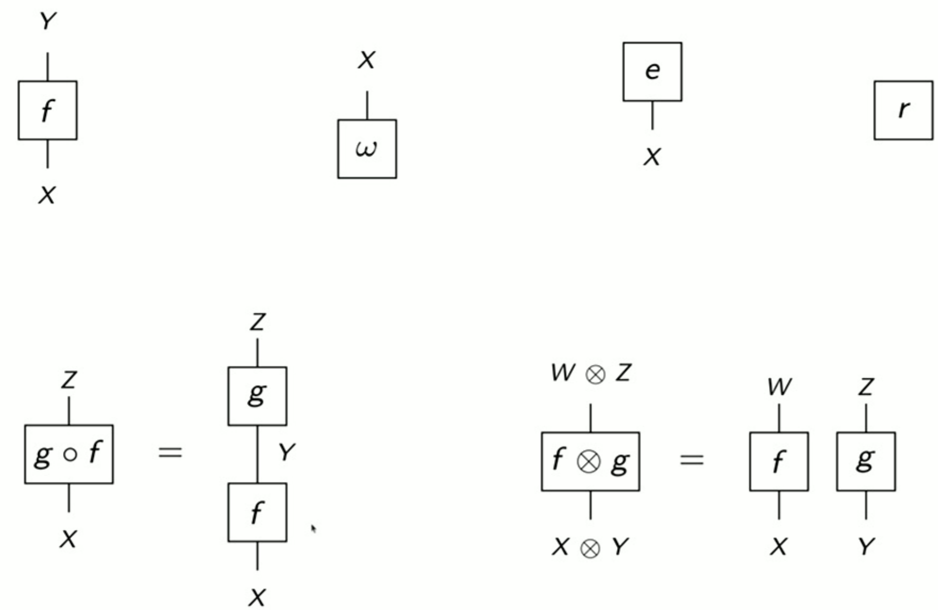
The *framework of causal models*, pioneered by Pearl and his collaborators, as well as Spirtes, Glymour

[arXiv:2304.07638]

Contents

Many aspects of causal reasoning are
most naturally and intuitively done as diagrammatic reasoning.

A process theory

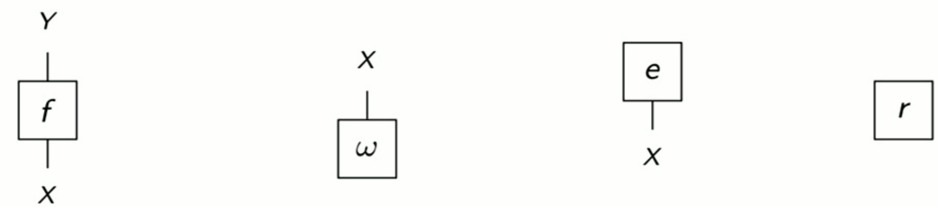


(subject to various axioms)

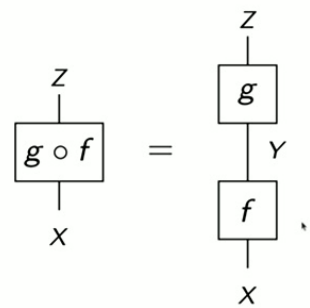
Symmetric monoidal category **C**.

A process theory

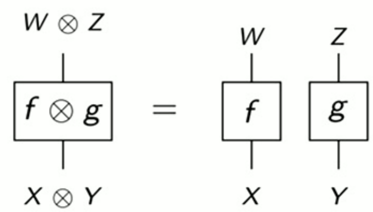
$\mathbf{Mat}_{\mathbb{R}^+}$



$(x, y) \mapsto f(y \mid x) \in \mathbb{R}^+$ $\quad x \mapsto \omega(x)$ $\quad x \mapsto e(x)$ $\quad r \in \mathbb{R}^+$



matrix multiplication

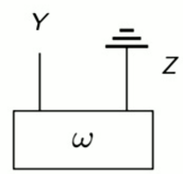


tensor product

(subject to various axioms)

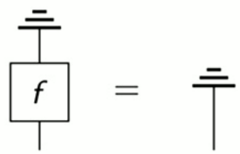
Symmetric monoidal category **C**.

What discard gives us

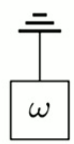


$:: y \mapsto \sum_{z \in Z} \omega(y, z)$

f is a **channel** iff



stochastic map *f*

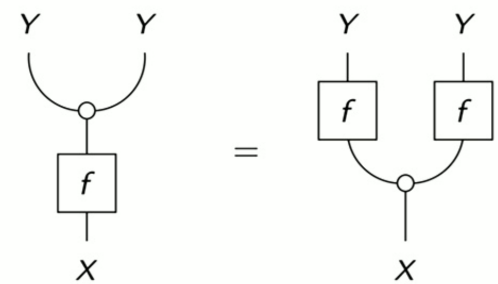


$= 1$

probability distribution ω

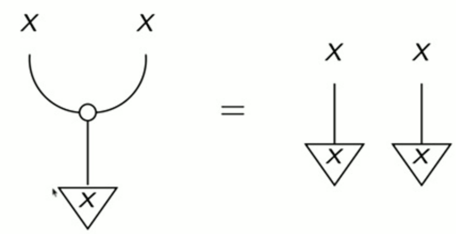
What copying gives us

Deterministic processes:

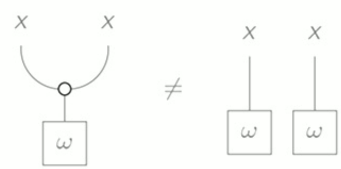


function
$X \to Y$

Sharp states:



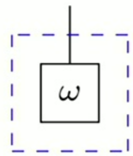
point
distribution



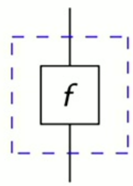
Markov category $\hookrightarrow$ cd-category

FStoch $\mathbf{Mat}_{\mathbb{R}^+}$

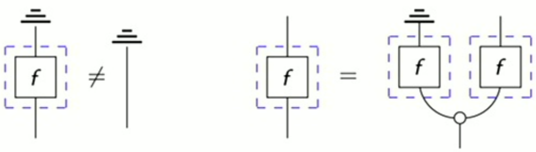
Further ingredient: normalisation



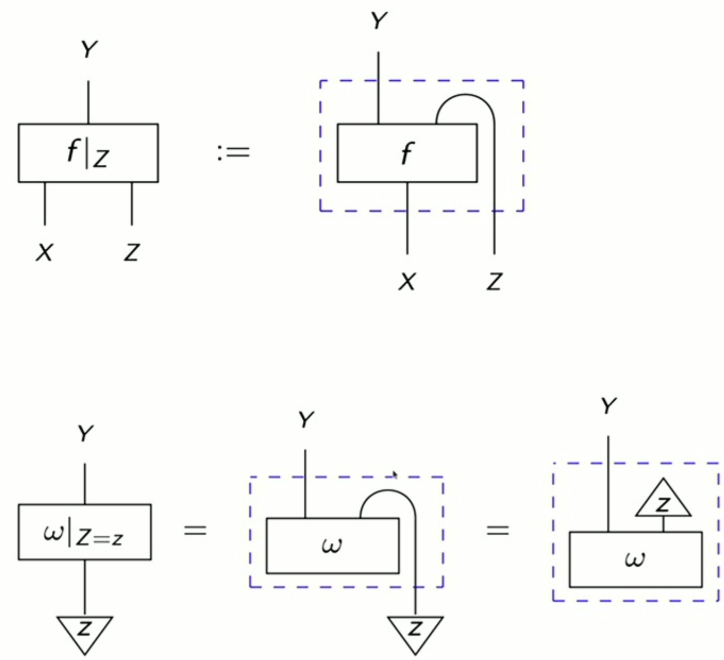
$$\mathrm{norm}(\omega)(x) = \frac{\omega(x)}{\sum_{x' \in X} \omega(x')}$$



Generally only a partial channel:



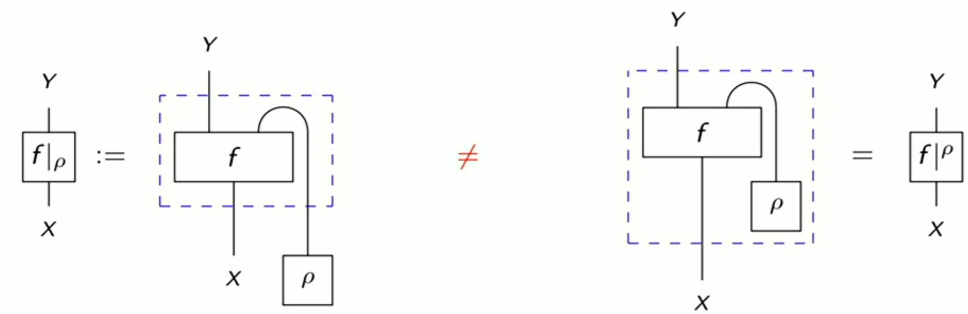
Conditionals



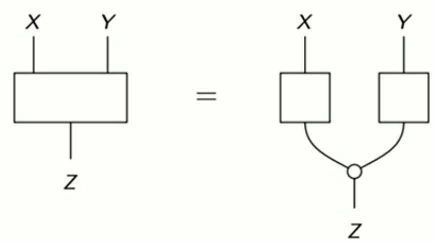
$P(Y|Z = z)$

Note that

Soft conditioning (fuzzy facts) is a non-trivial business – Jeffrey- vs Pearl-style:



Induced diagrammatic notion of **conditional independence**.



Part II

Causal models

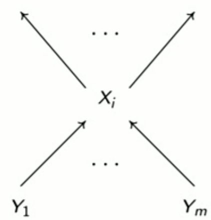
Syntax: causal structure

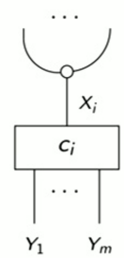
Causal structure (G, O)

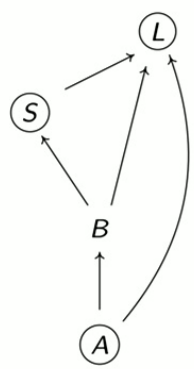
DAG G with vertices $V = \{X_1, \dots, X_n\}$,
output vertices $O \subseteq V$

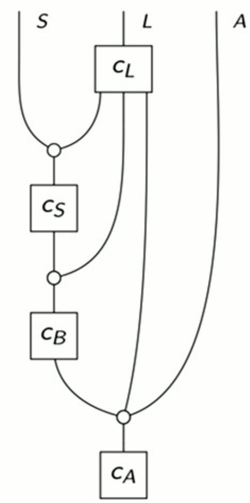
$\leftrightarrow$

(State) **network diagram**



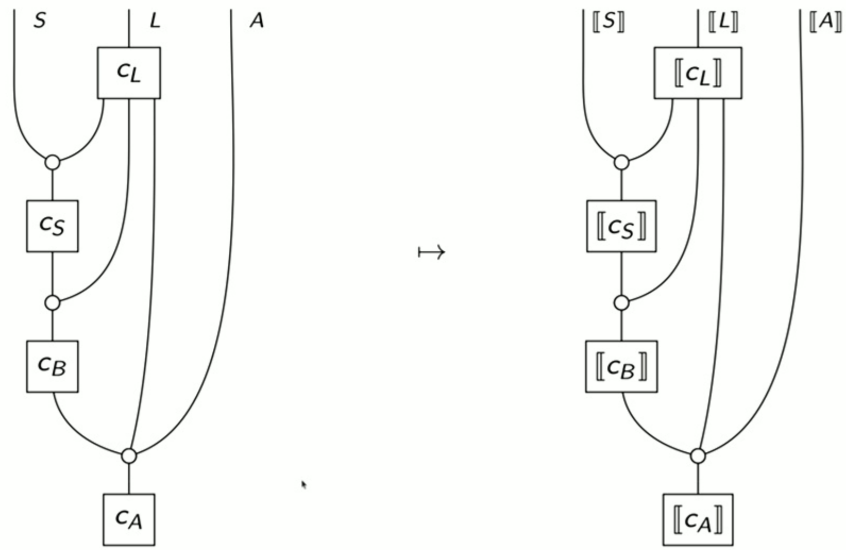






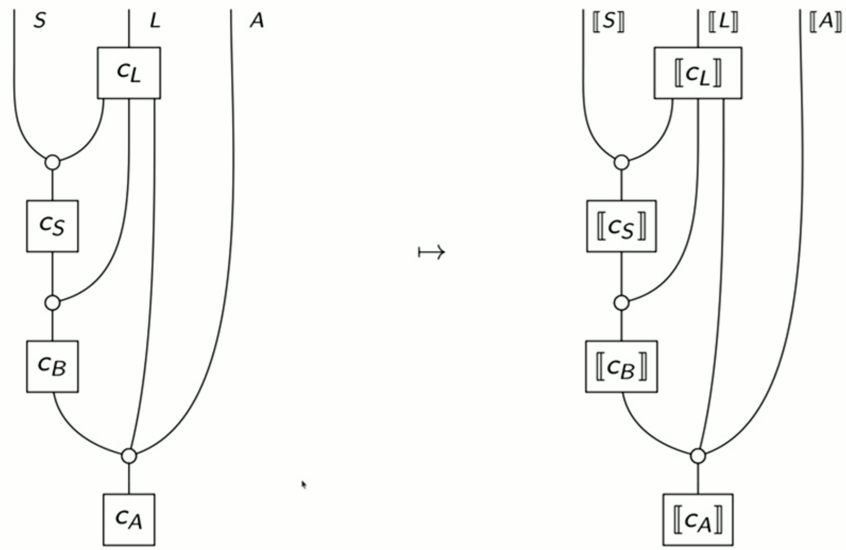
Semantics: causal model in $\mathbf{C}$

Definition: Let $\mathbf{C}$ be a cd-category. A *causal model* $\mathbb{M}$ in $\mathbf{C}$ is given by a state network diagram D with an interpretation $[\![-]\!]$ in $\mathbf{C}$



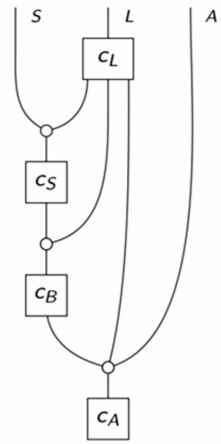
Semantics: causal model in **C**

Definition: Let **C** be a cd-category. A *causal model* $\mathbb{M}$ in **C** is given by a state network diagram D with an interpretation $[\![-]\!]$ in **C**

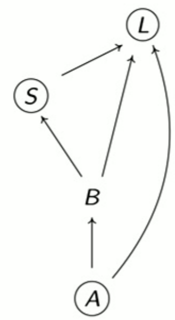


Equivalently: $[\![\ldots]\!] : \mathbf{Free}(D) \to \mathbf{C}$ [Jacobs, Kissinger and Zanasi 2019] , (G, O) $\{c_i : Pa(X_i) \to X_i\}_{X_i \in V}$, $\cdots$

A causal model in $\mathbf{C} = \mathbf{Mat}_{\mathbb{R}+}$ is a *causal Bayesian network*.



$\leftrightarrow$

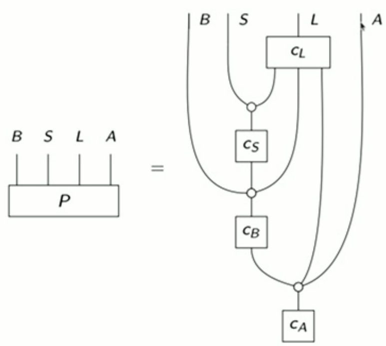


$P(L|SBA)$

$P(S|B)$

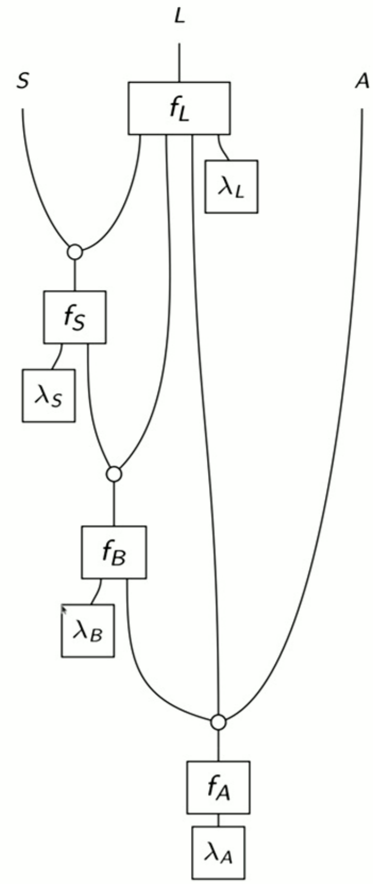
$P(B|A)$

$P(A)$

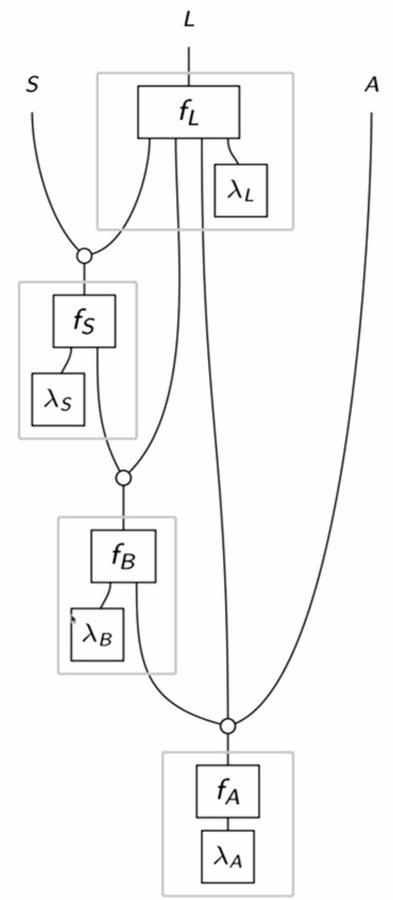


$$P(S, L, A) = P(L|SBA)\ P(S|B)\ P(B|A)\ P(A)$$

Functional causal models (SEMs)

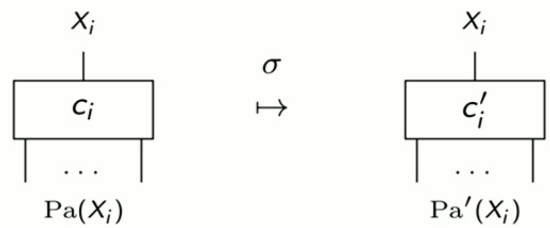


Functional causal models (SEMs)



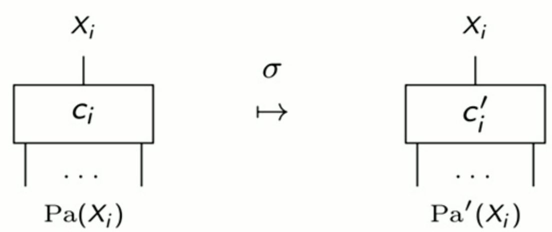
General interventions

An *intervention* σ on a causal model $\mathbb{M}$ in **C** is a modification of the mechanisms to yield a new causal model $\mathbb{M}'$ in **C** with the same variables.



General interventions

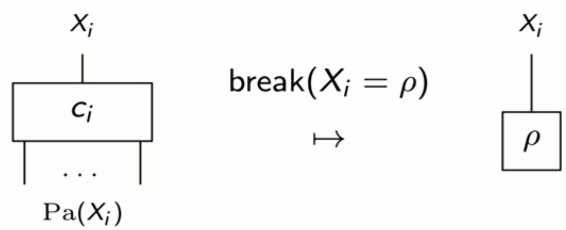
An *intervention* σ on a causal model $\mathbb{M}$ in **C** is a modification of the mechanisms to yield a new causal model $\mathbb{M}'$ in **C** with the same variables.



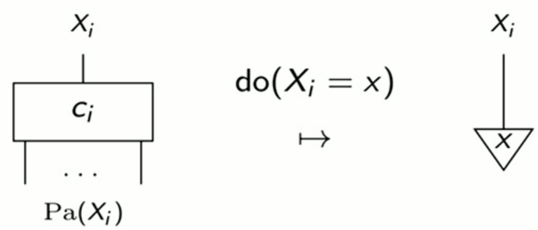
Write σ_X for $X \subseteq V$.

Zoo of prominent kinds

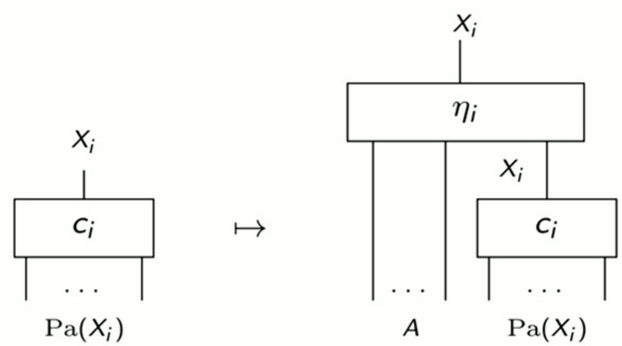
A *breaking intervention* at X_i:



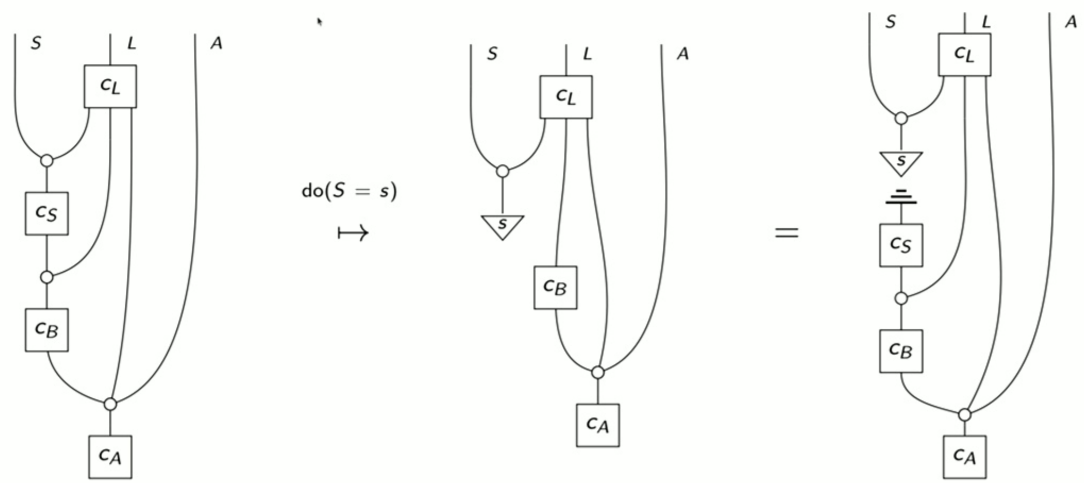
A *do intervention* $\mathrm{do}(X_i = x)$:



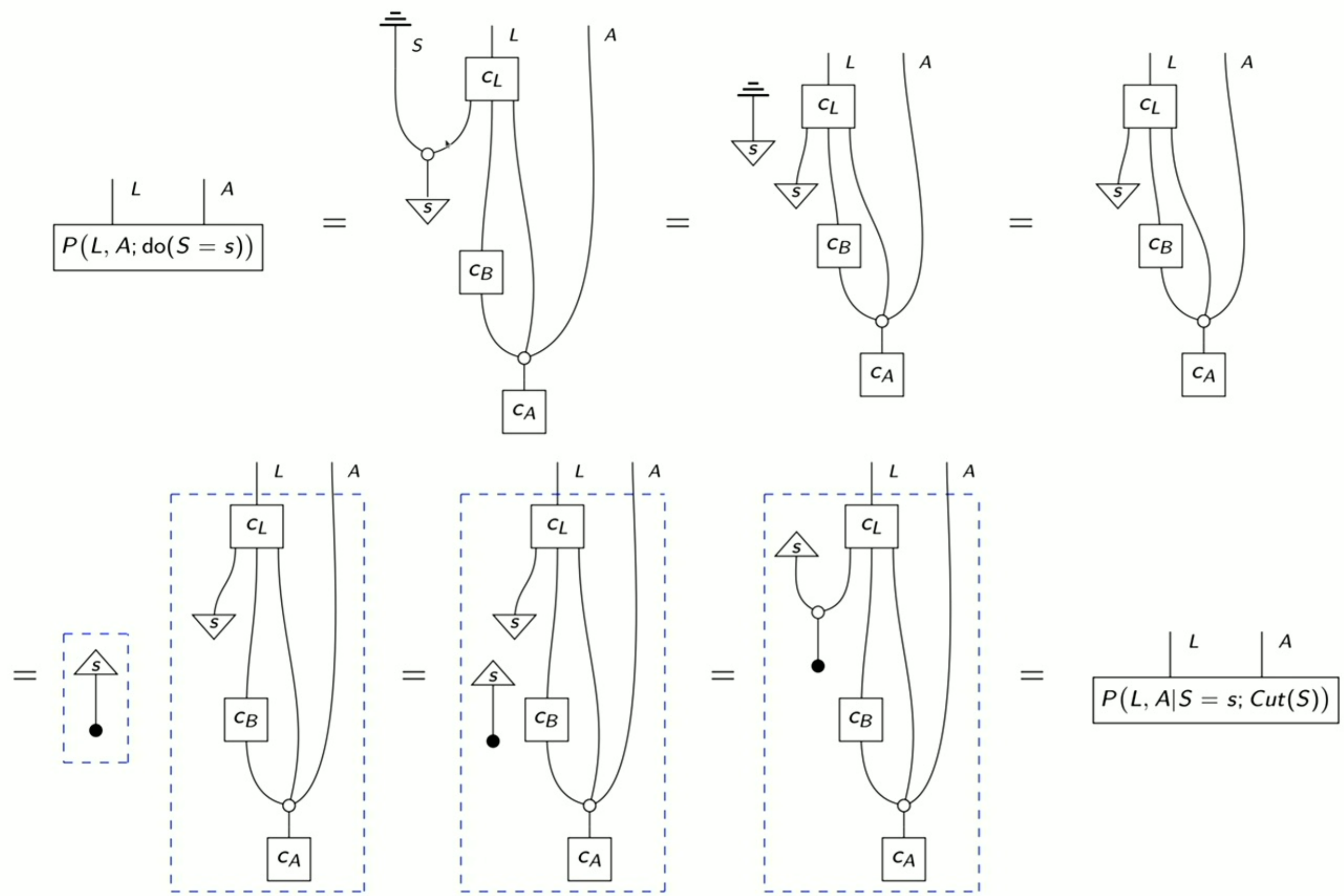
A *local intervention* at X_i:



Example do-intervention

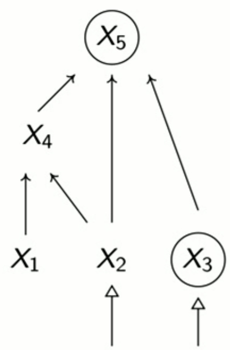


Example: do-intervention continued



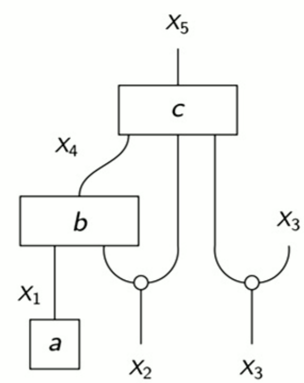
Open causal models

Open DAGs (G, I, O)

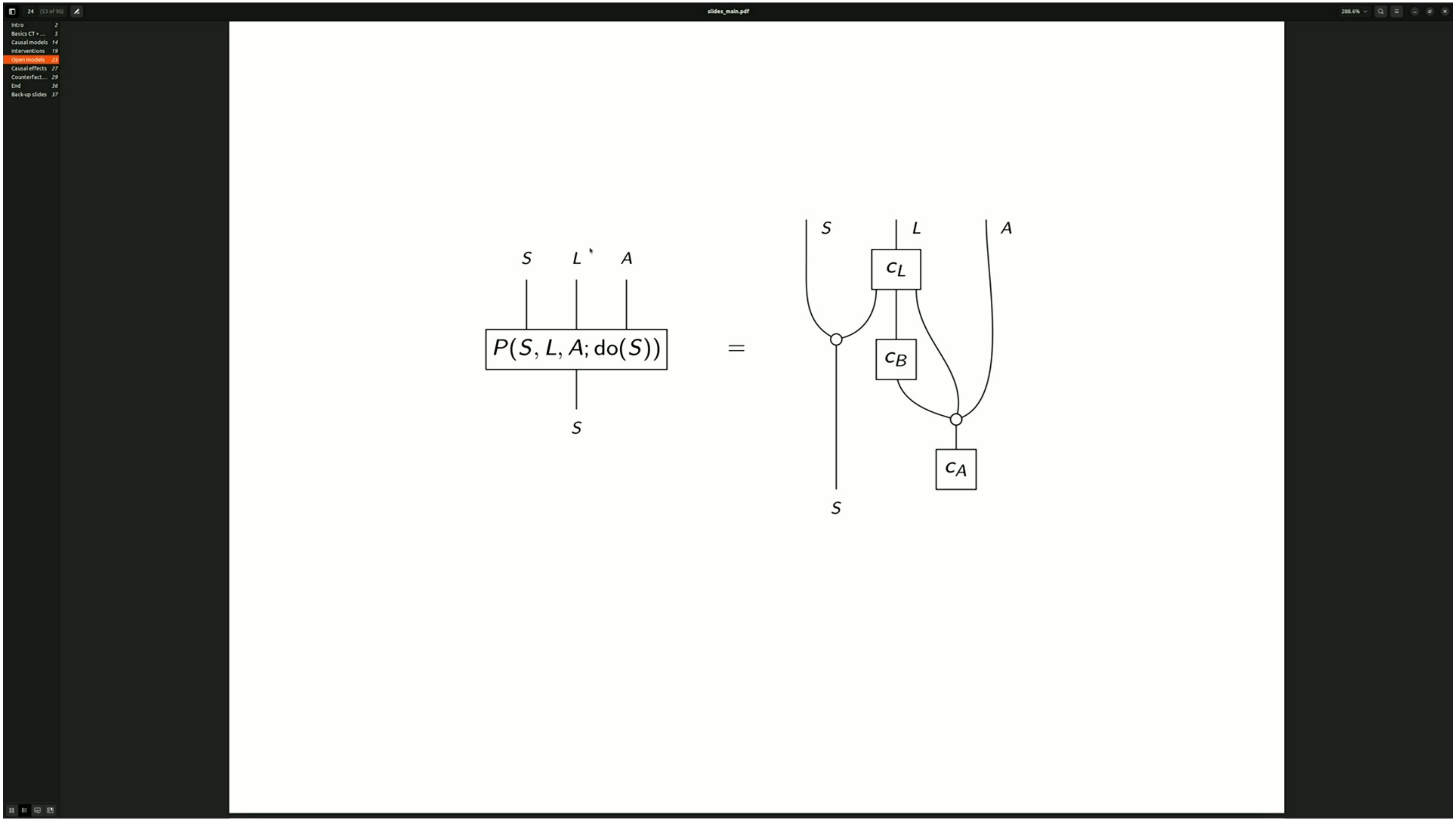


$\leftrightarrow$

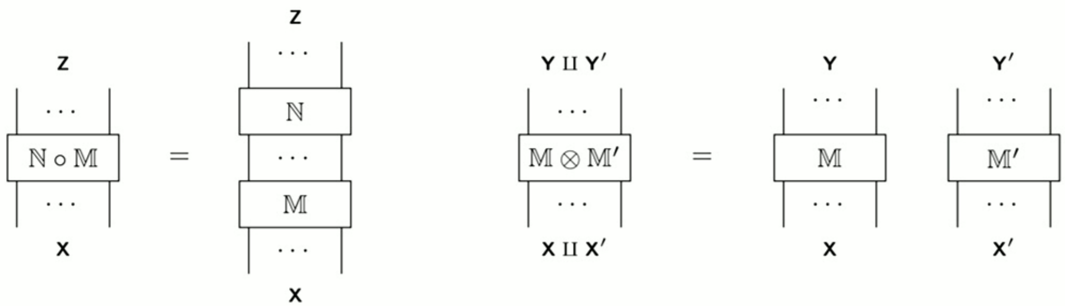
Network diagrams



Definition: An *open causal model* $\mathbb{M}$ in $\mathbf{C}$ is a network diagram D with interpretation $[\![-]\!]$ in $\mathbf{C}$.


slides_main.pdf
Intro 2
Basics CT + ... 5
Causal models 14
Interventions 19
Open models 23
Causal effects 27
Counterfact... 29
End 36
Back-up slides 37
S L A
P(S, L, A; do(S))
S
=
S L A
c_L
c_B
c_A
S

Categories of open causal models/open DAGs/network diagrams



Transformations of open causal models: opening, interventions, internalisations, externalisation. Probably also many others e.g. refinement and causal abstraction!

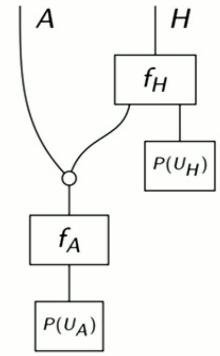
Part III

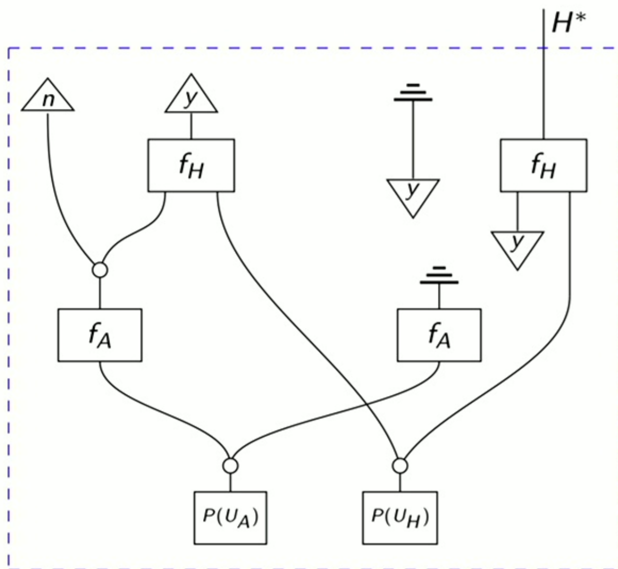
More examples, counterfactuals and problems of identifiability

slides_main.pdf

Counterfactuals

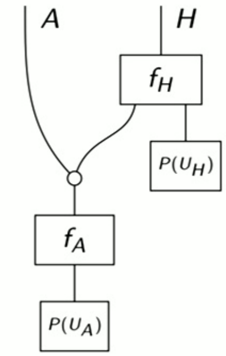
Today Mary woke up with a headache,
but had she taken an aspirin last night,
would she have woken up with a headache today?

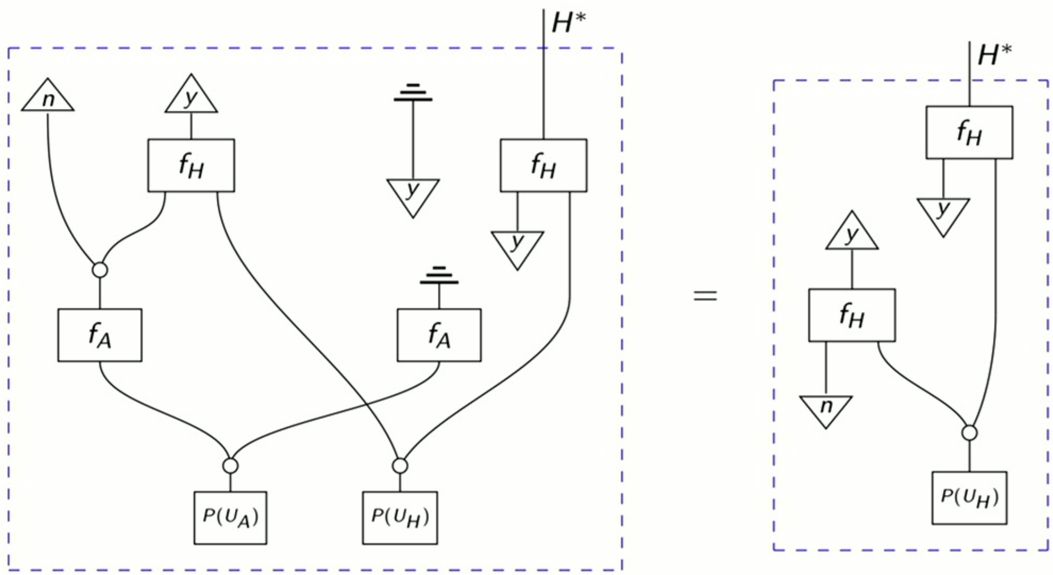


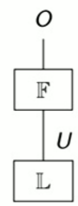


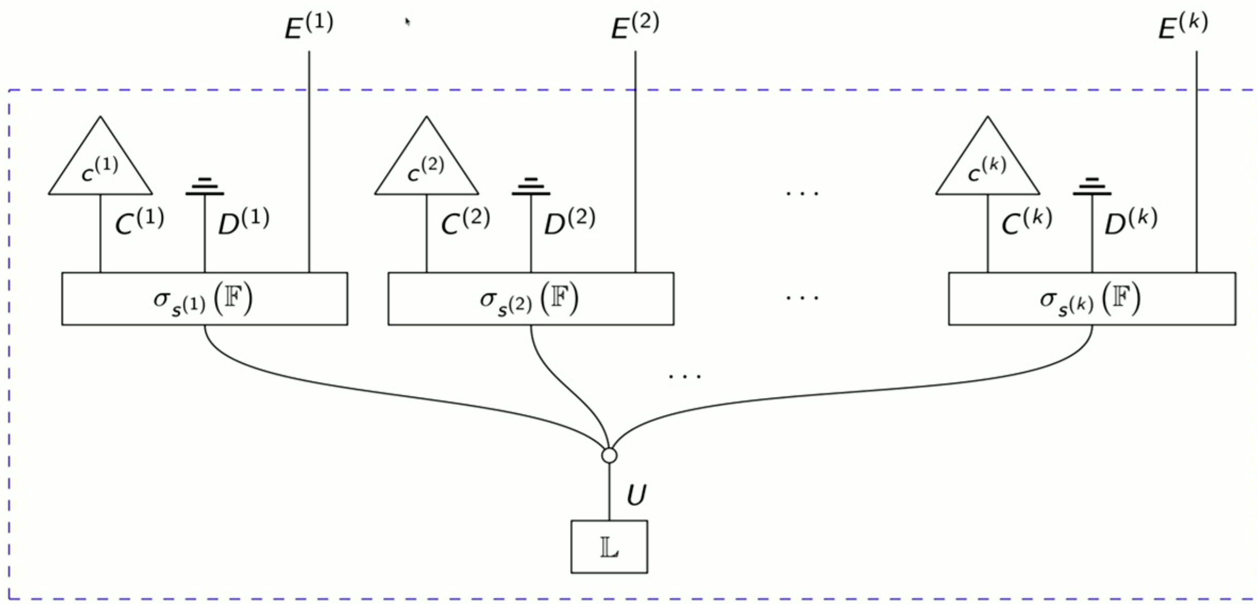
Counterfactuals

Today Mary woke up with a headache,
but had she taken an aspirin last night,
would she have woken up with a headache today?





Given functional causal model  in **C**, a *counterfactual* is a state of the form:



where $\sigma_{s^{(j)}}$ is a do-intervention $\text{do}(S^{(j)} = s^{(j)})$ and
for some j, j' with $j \neq j'$ it holds $C^{(j)} \neq \emptyset \neq E^{(j')}$.

Identifiability of counterfactuals

Given: Variables O and counterfactual terms $\left(s^{(j)}, c^{(j)}, E^{(j)}\right)_{j=1}^{k}$

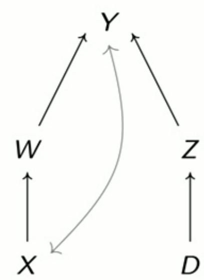
ADMG G with vertices O

$P_*(O)$

is the counterfactual uniquely determined?

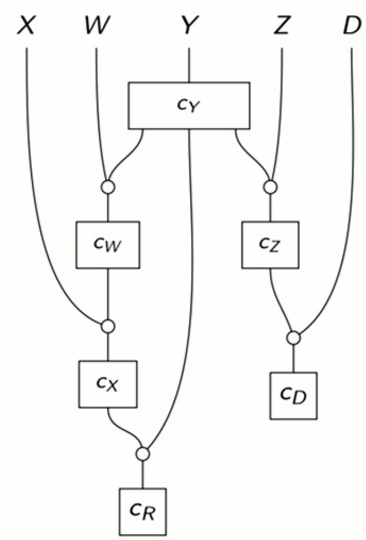
Example

Assuming



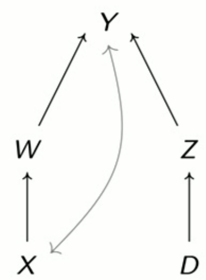
is this below counterfactual identifiable?

$$P_{\mathbb{M}_s^3}\left(Y^{(1)} \mid X^{(2)} = \tilde{x}, D^{(2)} = d, Z^{(3)} = z\right)$$



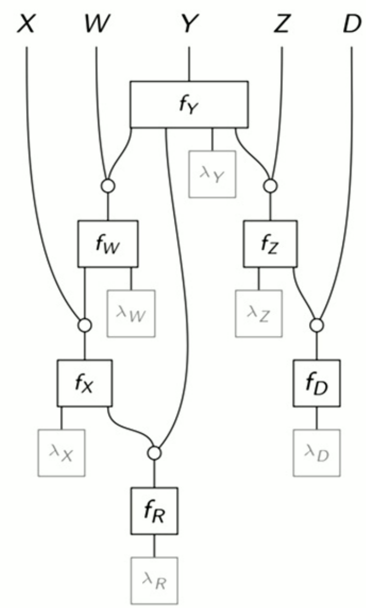
Example

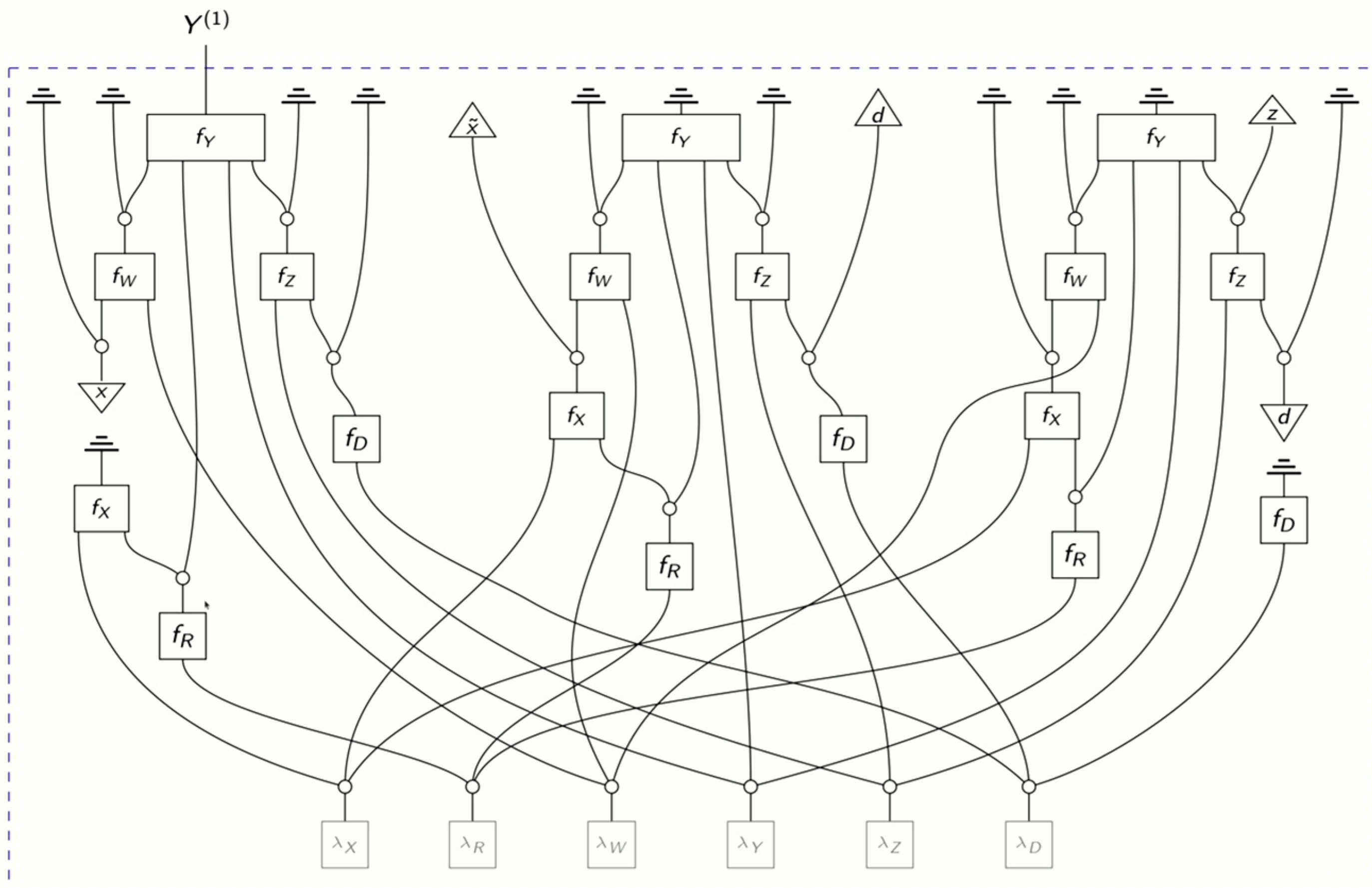
Assuming



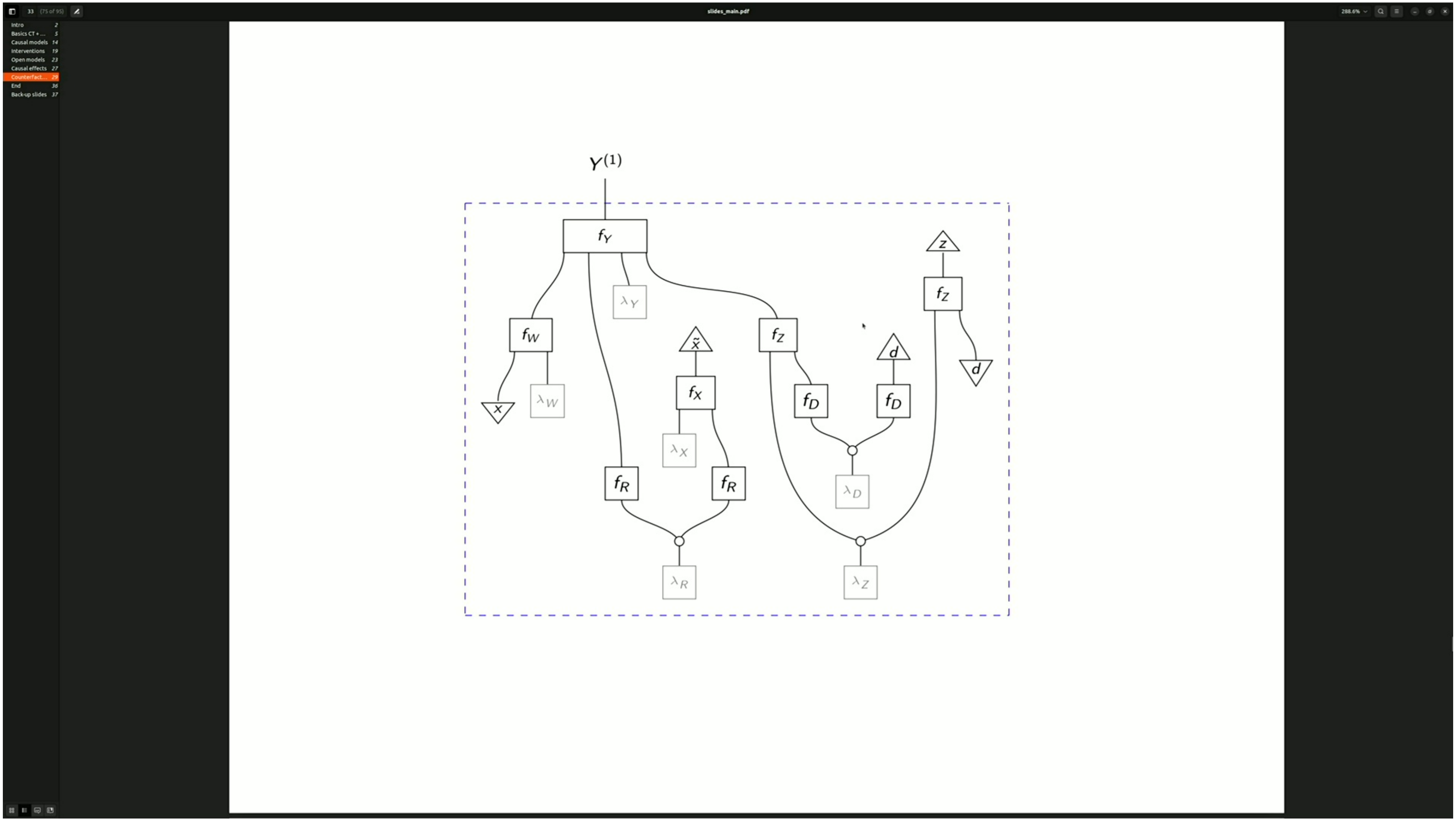
is this below counterfactual identifiable?

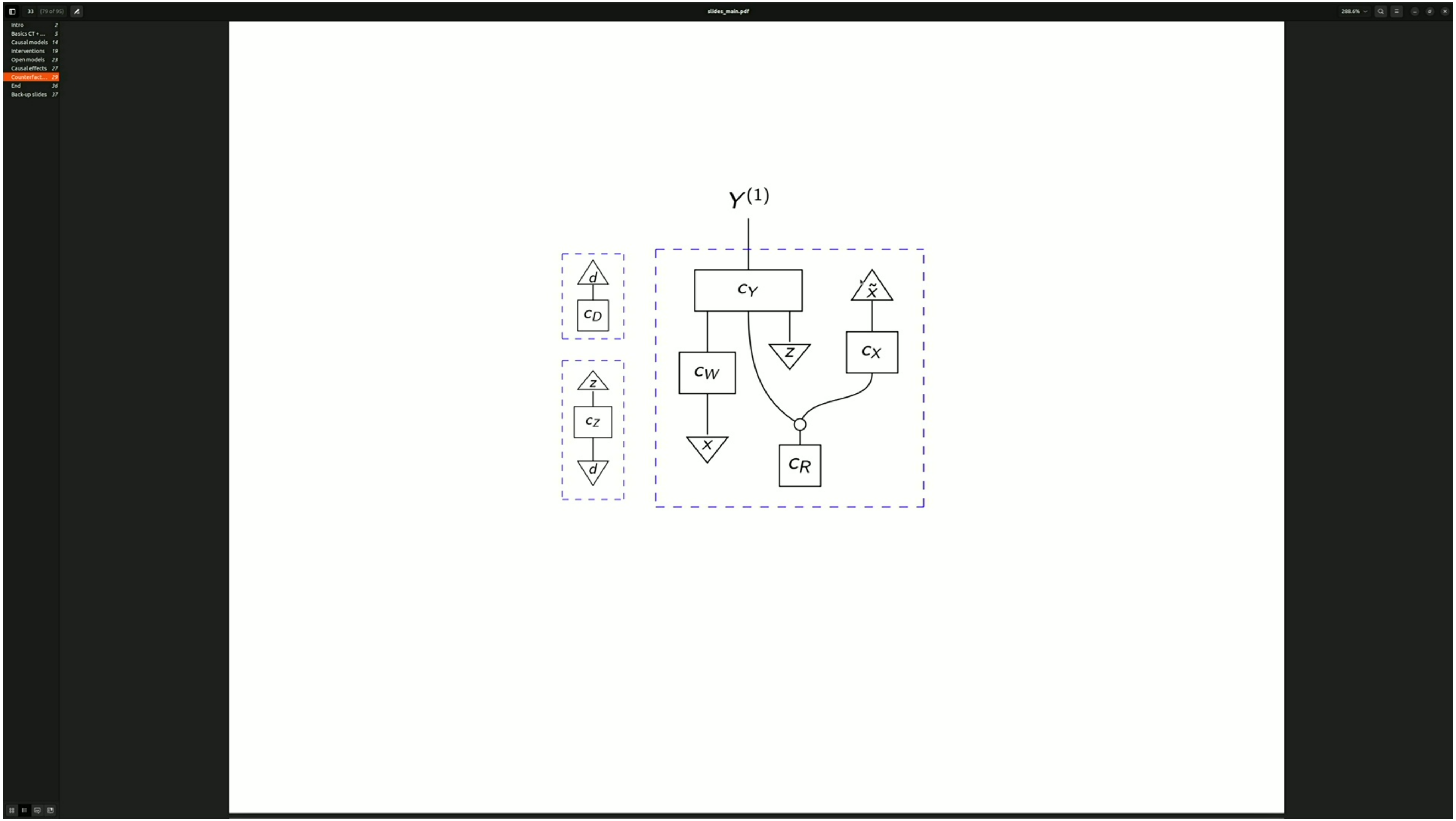
$$P_{\mathbb{M}_s^3}\left(Y^{(1)} \mid X^{(2)} = \tilde{x}, D^{(2)} = d, Z^{(3)} = z\right)$$

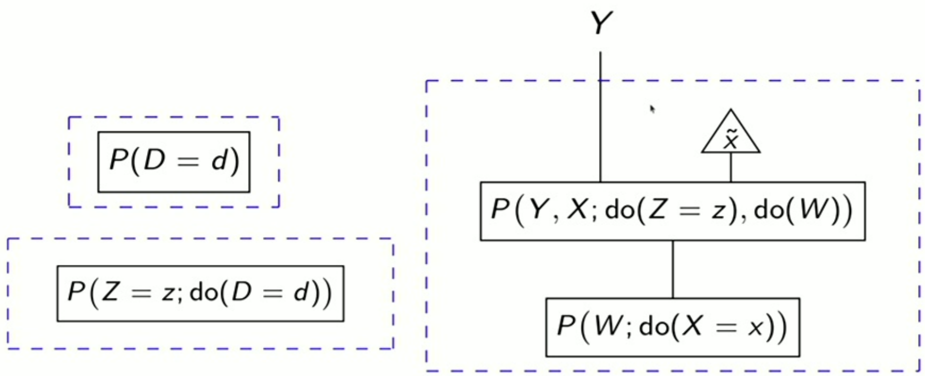




$Y^{(1)}$
f_Y
f_W
f_Z
f_X
f_D
f_R
x
$\tilde{x}$
d
z
λ_X
λ_R
λ_W
λ_Y
λ_Z
λ_D


slides_main.pdf
Intro 2
Basics CT + ... 5
Causal models 14
Interventions 19
Open models 23
Causal effects 27
Counterfact... 29
End 36
Back-up slides 37
$Y^{(1)}$
f_Y
λ_Y
f_W
λ_W
x
f_R
f_R
λ_R
$\tilde{x}$
f_X
λ_X
f_Z
f_D
f_D
d
λ_D
λ_Z
z
f_Z
d


slides_main.pdf
Intro 2
Basics CT + ... 5
Causal models 14
Interventions 19
Open models 23
Causal effects 27
Counterfact... 29
End 36
Back-up slides 37
$Y^{(1)}$
c_Y
$\tilde{x}$
z
c_X
c_W
x
c_R
d
c_D
z
c_Z
d


Y
$\tilde{x}$
$P(D = d)$
$P(Z = z; \mathrm{do}(D = d))$
$P(Y, X; \mathrm{do}(Z = z), \mathrm{do}(W))$
$P(W; \mathrm{do}(X = x))$

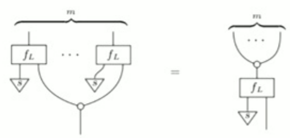
function `simplify-cf`(D, V, π):

INPUT: *Diagram D in $\mathbf{C}$ given by the network diagram of a parallel worlds model, possibly composed with some effects, the set V of variables of the functional model $\mathbb{M}$ that underlies the parallel worlds model and π a topological ordering*[34] *on V for the DAG $G_{\mathbb{M}}$.*

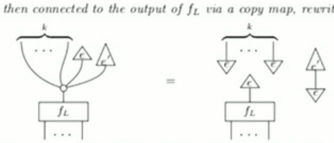
OUTPUT: *Simplified, but equivalent diagram D.*

1. *Let all discards 'fall through', i.e. make iterative use of the defining property of channels (Def. 3) and drop discarded wires wherever connected to a copy map (Def. 2) until no discards are left in the diagram.*
2. *Wherever a sharp effect is connected to a copy map use Eq. (6) to 'separate' all the involved wires from each other.*
3. *Starting with the lowest root node, iteratively go through the variables $L \in V$ in the order π and apply the below steps for the respective variable L.*

   (α) *Consider all those m appearances of the functional mechanism f_L from across the different worlds that share their inputs in the sense as on the left-hand side below and rewrite D accordingly:*



   (β) *If sharp effects are then connected to the output of f_L via a copy map, rewrite as:*



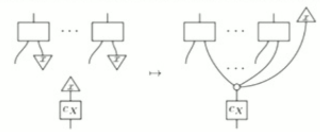
4. *For any $X \in V$ such that f_X appears in D with the state λ_X directly fed in, that is, rather than 'sharing' it with other occurrences of f_X via a copy map, replace it with c_X as in Eq. (53).*
5. *Output the rewritten diagram D.*

function `id-cf`$(G, W^k, P_*(O))$:

INPUT: *ADMG G with vertices O in $\mathbf{C}$, counterfactual terms $W^k = \left(s^{(j)}, c^{(j)}, E^{(j)}\right)_{j=1}^{k}$ and set $P_*(O)$.*

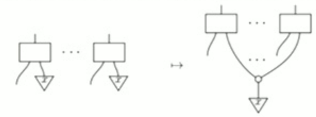
OUTPUT: *FAIL or the counterfactual C corresponding to W^k, assuming G, and expressed in terms of $P_*(O)$.*

1. *Let R be the set of additional root nodes introduced by the rootification $\tilde{\rho}(G)$. Let $\mathbb{M} = \mathbb{F} \circ \mathbb{L}$ be a corresponding FCM with endogenous variables $V := O \cup R$ and background variables $\{U_X\}_{X \in V}$ such that $\pi_O(G_{\mathbb{M}}) = G$ and such that it is compatible with P_*. For any $X \in V$ write c_X for the corresponding probabilistic mechanism obtained from feeding λ_X into f_X as in Eq. (53). Finally, let C be the counterfactual defined by W^k on the basis of $\mathbb{M}$ and let D be the same diagram up to normalisation, $C = \mathrm{norm}(D)$.*
2. $D =$ `simplify-cf`(D, V, π) for some topological order π for $\tilde{\rho}(G)$.
3. *If $\exists X \in V$ s.th. λ_X was not absorbed into c_X by* `simplify-cf`, *output* FAIL, *otherwise continue.*
4. *For $F_l \in \{F_1, ..., F_m\}$, the set of R-fragments of D:*
   1. *For each $X \in O$ such that X appears as the type of a wire in F_l:*
      1. *If mechanism c_X is a component of F_l:*
         a. *If the output wire of c_X does not have some sharp effect x on it, i.e. is fed into some other R-fragment, or is an output of D, and no other X type wire appears in F_l, do nothing.*
         b. *If the output wire of c_X is composed with some sharp effect x in D and any other wires of type X in F_l all have the state x fed into them, rewrite D according to:*



(58)

         c. *Else output* FAIL.
      2. *If mechanism c_X is not a component of F_l:*
         a. *If all X type wires, input to F_l, are connected via copy maps to the same output of c_X in some other R-fragment, do nothing.*
         b. *If all X type wires, input to F_l, are fed the same state x, rewrite D according to:*



(59)

         c. *Else output* FAIL.
   2. *Replace the thus rewritten R-fragment $\widetilde{F}_l$ – same as F_l up to more copy maps – according to:*

$$\boxed{P(C_l, F_l^{out} ; \mathrm{do}(D_l), \mathrm{do}(F_l^{in}))} \quad \text{outputs } C_l,\ F_l^{out};\ \text{inputs } D_l,\ F_l^{in} \qquad (60)$$

   where D_l, F_l^{in}, C_l and F_l^{out} are the sets of objects of wires going into $\widetilde{F}_l$ with sharp states fed into them, the remaining inputs to $\widetilde{F}_l$, wires coming out of $\widetilde{F}_l$ with sharp effects on them, and the remaining outputs of $\widetilde{F}_l$, respectively.
5. *Output* $\mathrm{norm}(D)$.

Essencially analogous to `make-cg` and `IDC`* [Shpitser, Pearl (2008)].

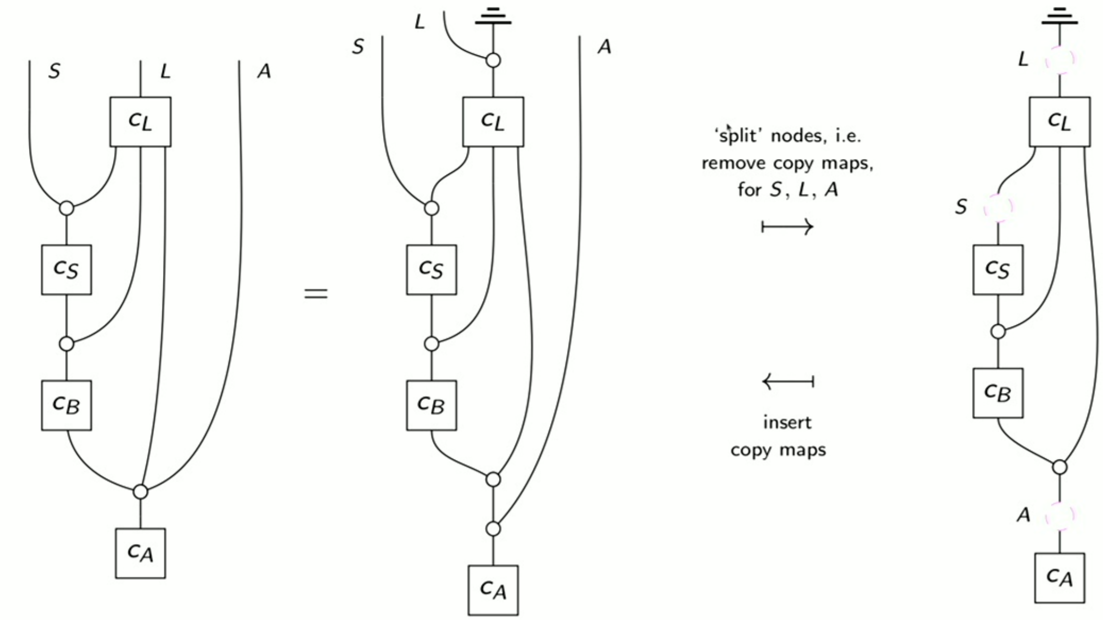
Though there is much yet to explore...

- **Conditioning:** diagrammatic axioms for more general measure theoretic set-up
- **Open causal models:** interesting categorical structure?

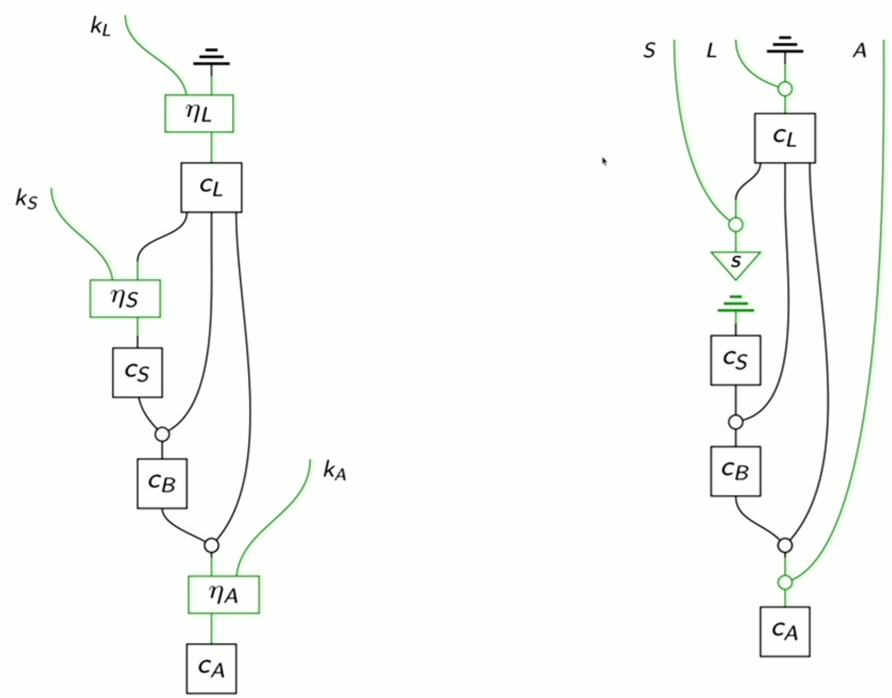
- **Causal effect identifiability:** fully general diagrammatisation of known results!
- **Counterfactuals:** identifiability with general interventions
- **Fuzzy updating:** soft conditioning and identifiability problems
- **Latent structure:** diagrammatic representation (that is, directly, without rootification)?
- **Cyclic causal models:** diagrammatic representation (and contrast with quantum foundations)
- **Higher-order map:** when is this view superior (for classical causal inference problems)?

- **Causal abstraction:** clarification and use in ML context?
- **XAI:** more generally clarification of causal aspects of interpretability
  (see [Tull, RL, Clark, Khan & Coecke, *Towards compositional interpretability for XAI*, 2024].)
- **Causal representation learning**?

Higher-order map view: split-node models

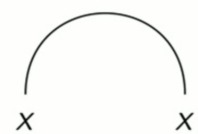


Higher-order map view: split-node models



Another ingredient: bending wires

A cd-category **C** has *caps* if $\forall X \; \exists$ below effect (subject to certain axioms):



$(x, y) \mapsto \delta_{x,y}$