Title: A Semantics for Counterfactuals in Quantum Causal Models

Speakers: Eric Cavalcanti

Series: Quantum Foundations, Quantum Information

Date: September 18, 2024 - 4:35 PM

URL: https://pirsa.org/24090122

Griffith UNIVERSITY
Centre for Quantum Dynamics

A semantics for counterfactuals in quantum causal models

Eric G. Cavalcanti

Joint work with Ardra Kooderi Suresh and Markus Frembs

arXiv:2302.11783v2

Causalworlds
Perimeter Institute, 18 September, 2024

Australian Government
Australian Research Council

FQXi
FOUNDATIONAL QUESTIONS INSTITUTE

Similarity analysis of counterfactuals

- Lewis [2] “Analysis 2” (see also Stalnaker [1]):

“A counterfactual

‘If it were that A, then it would be that C’

is (non-vacuously) true if and only if some (accessible) world where both A and C are true is more similar to our actual world, overall, than is any world where A is true but C is false”.

[1] R. C. Stalnaker, “A Theory of Conditionals”, in *Studies in Logical Theory*, N. Rescher (ed.), 98–112 (1968).
[2] D. K. Lewis, “Counterfactuals and Comparative Possibility”, *Journal of Philosophical Logic*, 2(4): 418–446 (1973).
[3] D. K. Lewis, “Counterfactual Dependence and Time’s Arrow”, *Noûs*, 13(4): 455–476 (1979)

“Unperformed experiments have no results”

“No Counterfactual Definiteness”

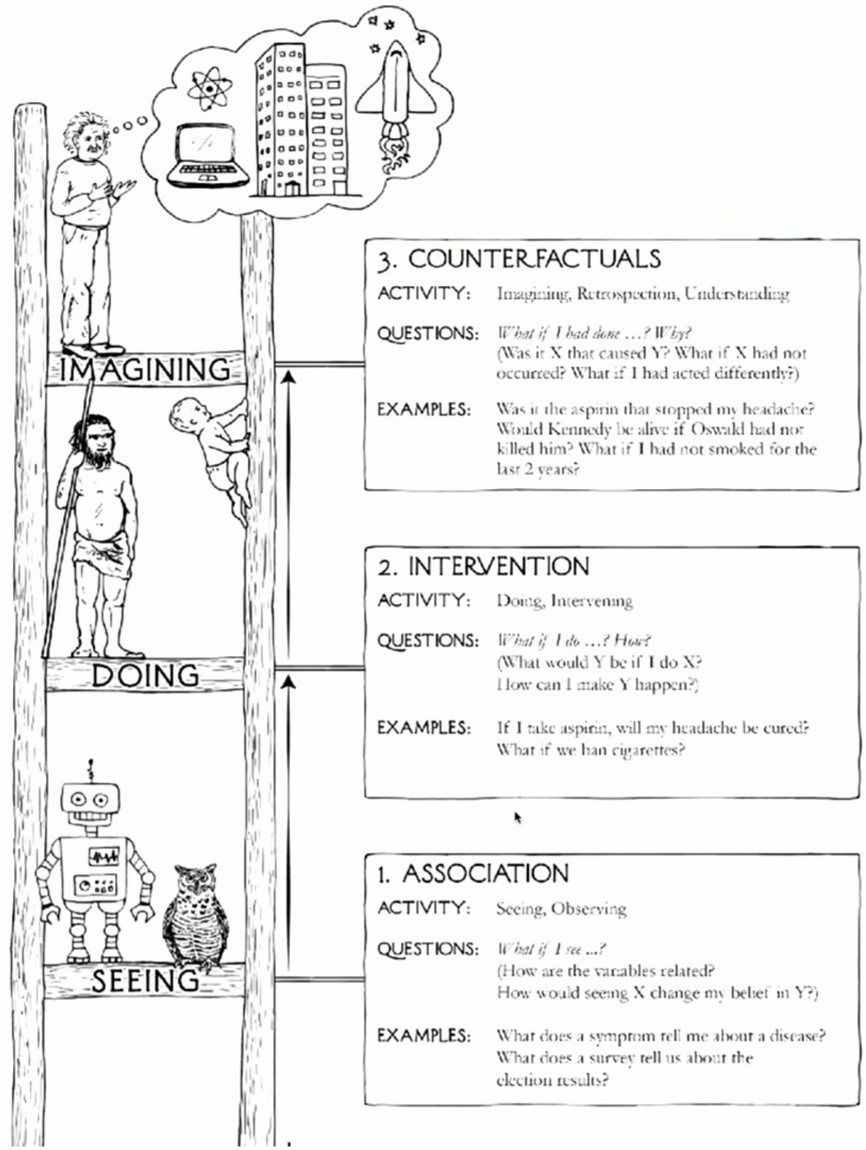
2

Counterfactuals and causation

- Lewis [5]: counterfactual analysis of causation
  - assumes determinism, as does his theory of counterfactuals
- Pearl [6]: causal analysis of counterfactuals
  - also assumes determinism (structural causal model)
- For both: counterfactual dependence implies causal dependence

[5] D.K. Lewis, "Causation", *Journal of Philosophy*, 70(17): 556–567 (1973)
[6] J. Pearl. Causality: Models, Reasoning and Inference. Cambridge University Press, (2000).

Pearl's ladder of causation



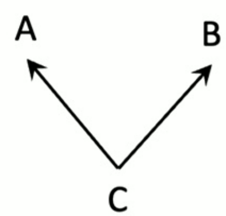
Level 1 - Association

- Bayesian networks
  - Directed Acyclic Graph (DAG) *G* over a set of vertices **V =** $\{V_1, \ldots, V_n\}$

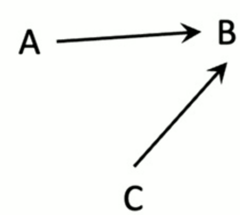
Markov condition: $P(V_1, \ldots, V_n) = \prod_i P(V_i|Pa(V_i))$

  - E.g.:

P(A,B,C) = P(A|C)P(B|C)P(C)



P(A,B,C) = P(B|A,C)P(A)P(C)

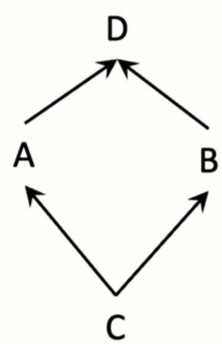


Level 2 - Intervention

- Causal Bayesian networks
  - Arrows represent causal relations

Def: *A Classical causal model* consists of a DAG G encoding a causal Bayesian network, and a probability distribution P that is Markov with respect to G.

G:



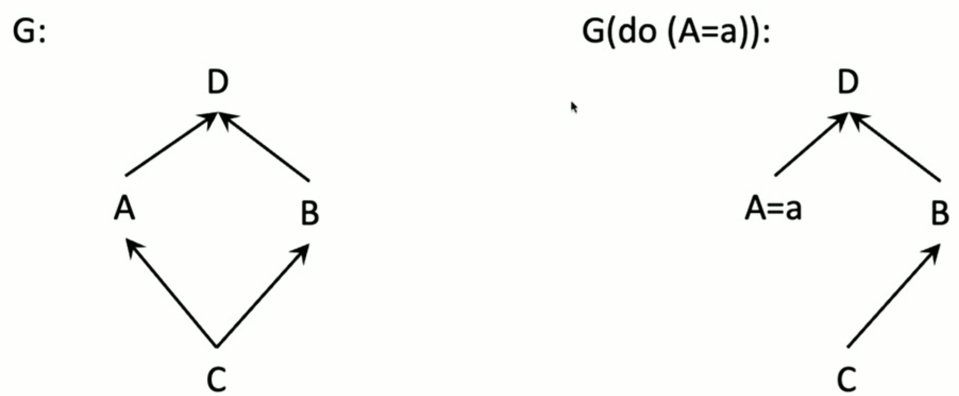
$$P(A,B,C,D) = P(D|A,B)P(A|C)P(B|C)P(C)$$

Level 2 - Intervention

– Oracle for interventions

do-intervention:

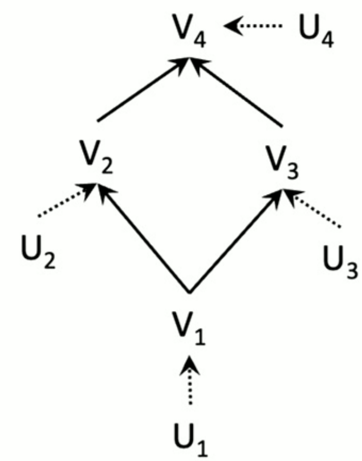
$$P_{do(A=a)}(A,B,C,D) = P(D|A=a,B)P(B|C)P(C)$$



Level 3 - Counterfactuals

- (Probabilistic) Structural Causal Model

E.g. $v_4 = f_4(v_2, v_3, u_4)$



  - $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F}, P \rangle$
  - Endogenous variables **V**=$\{V_1,\ldots,V_n\}$
  - Exogenous variables **U** = $\{U_1,\ldots,U_n\}$
  - Functions $\boldsymbol{F} = \{f_i \mid v_i = f_i(pa_i, u_i)\}_i$
  - Probability P(u)

→ Induces a CCM: a directed graph G (which we will restrict to DAGs) and a P(**V**) Markov to G.

Level 3 - Counterfactuals

- Interventions in a Structural Causal Model

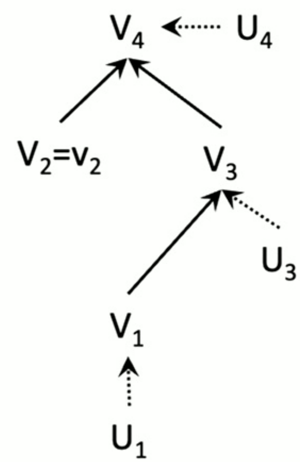
$M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$

Intervention $do(\boldsymbol{X} = \boldsymbol{x})$ on $\boldsymbol{X} \subset \boldsymbol{V}$ leads to a submodel

$M_x = \langle \boldsymbol{U_x}, (\boldsymbol{V_x}, \boldsymbol{X} = \boldsymbol{x}), \boldsymbol{F_x} \rangle,$

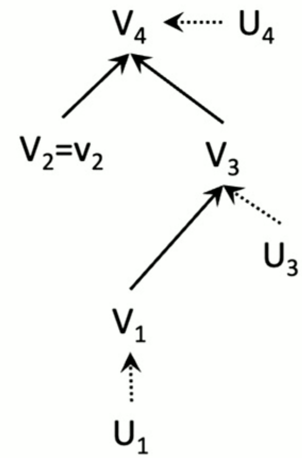
where $\boldsymbol{V_x} = \boldsymbol{V} \backslash \boldsymbol{X}$, $\boldsymbol{U_x} = \boldsymbol{U} \backslash \boldsymbol{U}(\boldsymbol{X})$, $\boldsymbol{F_x} = \boldsymbol{F} \backslash \boldsymbol{F}(\boldsymbol{X})$

E.g. do(V_2=v_2)



Level 3 - Counterfactuals

$M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$, $\boldsymbol{X}, \boldsymbol{Y} \subseteq \boldsymbol{V}$



Counterfactual proposition $\boldsymbol{Y}_{\boldsymbol{x}}(\boldsymbol{u}) = \boldsymbol{y}$

``**Y** would have been **y**, had **X** been **x**, in a situation specified by the background variables **U=u**"

X=x: antecedent

Y=y: consequent

$\boldsymbol{Y}_{\boldsymbol{x}}(\boldsymbol{u})$: potential response of **Y** to the action do(**X=x**) -- solution for **Y** in the submodel $M_{\boldsymbol{x}}$

→ do(**X=x**): "minimal surgery" required to make antecedent true

Level 3 - Counterfactuals

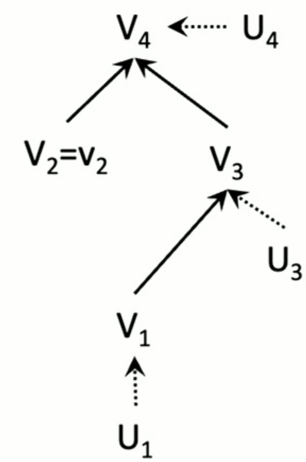
$\boldsymbol{M = \langle U, V, F \rangle}$, $\boldsymbol{X, Y \subseteq V}$

Counterfactual proposition $\boldsymbol{Y_x(u) = y}$

``**Y** would have been **y**, had **X** been **x**, in a situation specified by the background variables **U=u**"

Probability of the counterfactual $\boldsymbol{Y_x = y}$:

$$P(\mathbf{Y_x} = \mathbf{y}) = \sum_{\mathbf{u}|\mathbf{Y_x}(\mathbf{u})=\mathbf{y}} P(\mathbf{u})$$



Pearl's three-step algorithm for counterfactuals

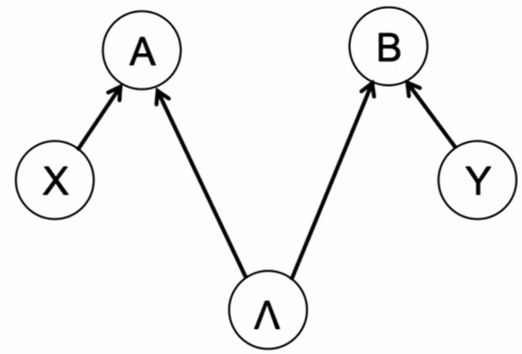
Given evidence e, what is the probability that Y would have been y had X been x?

1. Abduction
   - Update $P(u)$ by the evidence *e* to obtain $P(u|e)$
2. Action
   - Replace the equations for the variables in *X* by *X*=*x*. (do(*X*=*x*))
3. Prediction
   - Compute the probability $P(Y = y)$ using the modified model $\langle M_x | P(u|e) \rangle$.

J. Pearl, *Causality: Models, Reasoning and Inference,* Cambridge Univ. Press (2009)

Quantum violations of CCMs

– Bell's theorem
  - Classical causal models + relativistic causal structure (+ exogenous interventions)
    →contradiction with quantum correlations

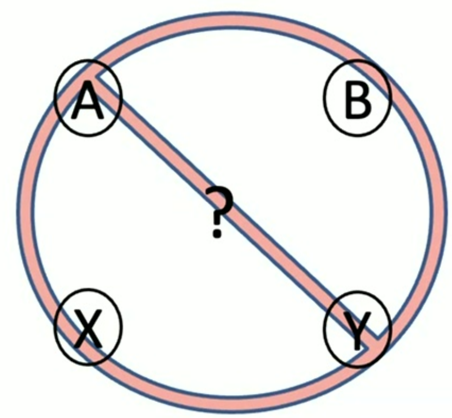


– Fine-tuning theorems
  - No classical causal model can explain violations of Bell or Kochen-Specker inequalities without fine-tuning

Wood and Spekkens, NJP **17**, 33002 (2015)

EGC, Phys. Rev. X 8, 021018 (2018)

J.C Pearl and EGC, Quantum 5, 518 (2021)



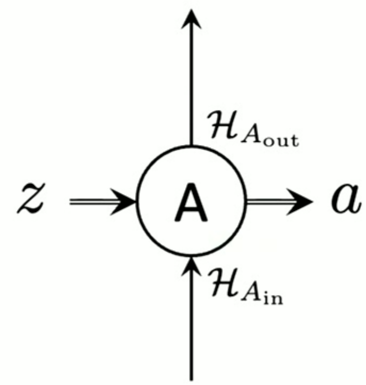
→ Quantum Causal Models

Quantum causal models

- Quantum node: locus of *potential interventions*
  - Input/output Hilbert spaces, with $d_{A_{in}} = d_{A_{out}}$
- Quantum instrument:
  - Set of completely positive (CP) maps $\mathcal{M}^a$ summing to a completely positive trace-preserving (CPTP) map $\mathcal{M}_A$.

$$\mathcal{I}_A^z = \{\mathcal{M}_A^{a|z} : \mathcal{L}(\mathcal{H}_{A^{\text{in}}}) \to \mathcal{L}(\mathcal{H}_{A^{\text{out}}})\}_a$$

$$\mathcal{M}_A = \sum_a \mathcal{M}_A^{a|z}$$



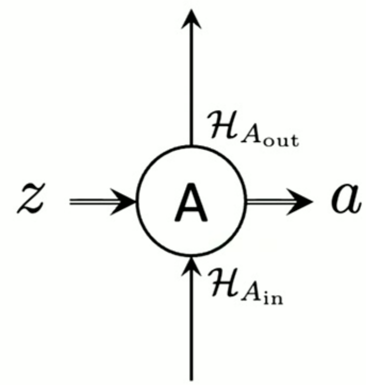
Costa, Shrapnel, New J. Phys. 18 063032 (2016), arXiv:1512.07106
Barrett, Lorenz, Oreshkov, arXiv:1906.10726

Quantum causal models

- Choi-Jamiolkowski isomorphism for instruments

$$\tau_A^{a|z} = \sum_{i,j} \mathcal{M}_A^{a|z}(|i\rangle\langle j|)^T_{A^{\mathrm{out}}} \otimes |j\rangle\langle i|_{A^{\mathrm{in}}}$$

$$\tau_A^{|z} = \sum_a \tau_A^{a|z}$$



- Trace preservation of $\mathcal{M}_A = \sum_a \mathcal{M}_A^{a|z}$ implies:

$$\mathrm{Tr}_{A^{\mathrm{out}}}[\tau_A^{|z}] = \mathbb{I}_{A^{\mathrm{in}}}$$

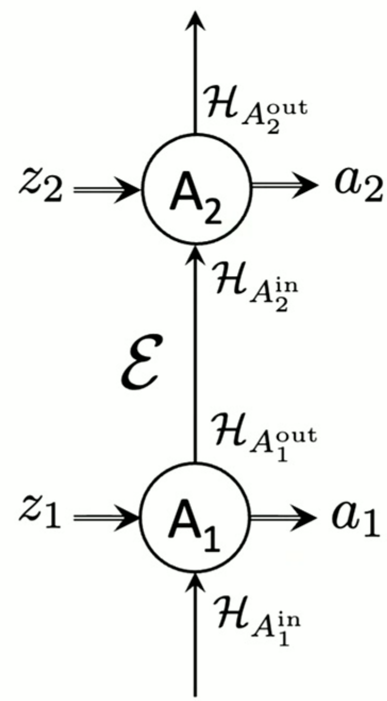
Quantum causal models

- CJ isomorphism for channels (CPTP maps) between nodes:

$$\mathcal{E} : \mathcal{L}(\mathcal{H}_{A_1^{\text{out}}}) \to \mathcal{L}(\mathcal{H}_{A_2^{\text{in}}})$$

$$\rho_{A_2|A_1} = \sum_{i,j} \mathcal{E}(|i\rangle\langle j|)_{A_2^{in}} \otimes |i\rangle_{A_1^{out}}\langle j|$$

(Compare to $\tau_A^{a|z} = \sum_{i,j} \mathcal{M}^{a|z}(|i\rangle\langle j|)^T_{A_{out}} \otimes |j\rangle_{A_{in}}\langle i|$)



23

Quantum causal models

A ***quantum process operator**** $\boldsymbol{\sigma_{A_1,\ldots,A_n}}$ over nodes A_1, … , A_n is a positive semi-definite operator:

$$\sigma_{A_1,\cdots,A_n} \in \mathcal{L}(\bigotimes_i \mathcal{H}_{A_i^{in}} \otimes \mathcal{H}_{A_i^{out}})$$

such that for all choices of instruments z = (z_1 , … , z_n) at the nodes,

$$\mathrm{Tr}_{A_1\cdots A_n}[\sigma_{A_1,\cdots,A_n}\tau_{A_1}^{z_1} \otimes \cdots \otimes \tau_{A_n}^{z_n}] = 1$$

Probabilities are given by the *generalized Born rule*:

$$P(a_1,\cdots,a_n|z_1,\cdots,z_n) = \mathrm{Tr}_{A_1\cdots A_n}[\sigma_{A_1,\cdots,A_n}\tau_{A_1}^{a_1|z_1} \otimes \cdots \otimes \tau_{A_n}^{a_n|z_n}]$$

Where we use the shorthand notation:

$$\mathrm{Tr}_A[\cdots] = \mathrm{Tr}_{A_{in},A_{out}}[\cdots]$$

*Barrett, Lorenz, Oreshkov, arXiv:1906.10726

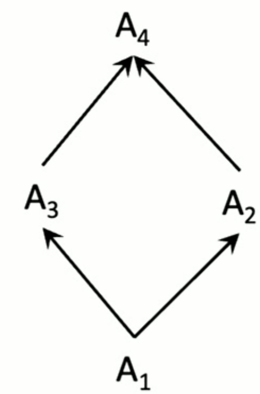
Quantum causal models

A ***quantum causal model**** is specified by

1) a DAG *G* with quantum nodes A_1 , … , A_n as vertices

2) A quantum channel $\rho_{A_i|Pa(A_i)}$ for each node such that

$$[\rho_{A_i|Pa(A_i)}, \rho_{A_j|Pa(A_j)}] = 0 \text{ for all } i, j.$$



A process operator given by

$$\sigma_{A_1,\cdots,A_n} = \prod_i \rho_{A_i|Pa(A_i)}.$$

This is the *quantum causal Markov condition.*

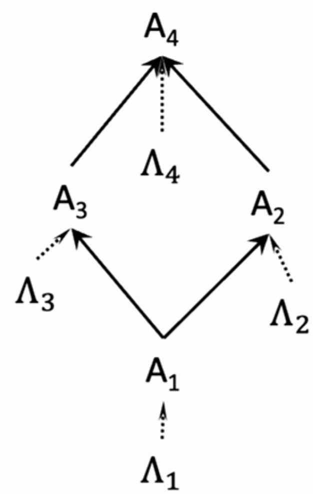
*Barrett, Lorenz, Oreshkov, arXiv:1906.10726

Quantum structural causal models

Classical SCM: full information about the causal relations, lack of information associated with events at exogenous nodes.

In a *quantum structural causal model (QSM),* lack of information about events at an exogenous node Λ is represented as a discard-and-prepare instrument

$$\{\tau_{\Lambda}^{\lambda}\}_{\lambda} := \{P(\lambda)(\rho_{\Lambda^{\mathrm{out}}}^{\lambda})^{T} \otimes \mathbb{I}_{\Lambda^{\mathrm{in}}}\}_{\lambda}$$



- Which instrument? *A: Effectively classical, well-decohered, "stable events"**
  - → *Allows Bayesian inference*

* Di Biagio and Rovelli, Foundations of Physics 51, 30 (2021)

Quantum structural causal models

- A ***quantum structural causal model*** M_Q is specified by:

1) A set of *quantum nodes* **A** = { A_1 , … , A_n}
2) A set of *exogenous nodes* $\mathbf{\Lambda} = \{\Lambda_1, \dots, \Lambda_n\}$
3) A *sink node S (not shown in figure)*
4) A *unitary channel* $\rho^U_{\boldsymbol{AS}|\boldsymbol{A\Lambda}}$ that satisfies the no-influence conditions

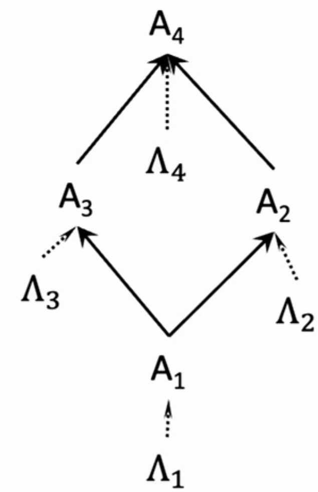
$$\{\Lambda_j \nrightarrow A_i\}_{j\neq i}$$

5) A set of *discard-and-prepare instruments* for every exogenous node

$$\{\tau^{\lambda_i}_{\Lambda_i}\}_{\lambda_i} \equiv \{p(\lambda_i)(\rho^{\lambda_i}_{\Lambda^{out}_i})^T \otimes \mathbb{I}_{\Lambda^{in}_i}\}_{\lambda_i}$$

- We also define (for later notational convenience)

$$\{\widetilde{\tau}^{\lambda}_{\Lambda}\}_\lambda := \{P(\lambda)(\rho^{\lambda}_{\Lambda^{\rm out}})^T \otimes \frac{1}{\dim(\mathcal{H}_{\Lambda^{\rm in}})}\mathbb{I}_{\Lambda^{\rm in}}\}_\lambda$$



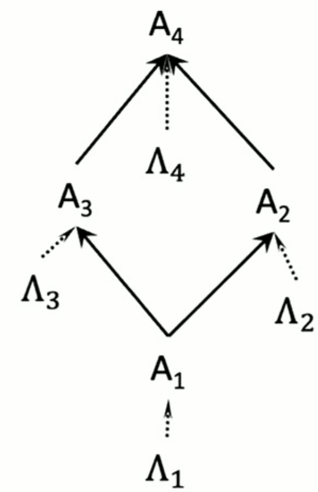
Quantum structural models

We say that a process operator $\sigma_{\boldsymbol{A}}$ is ***structurally compatible**** with a DAG G iff there exists a quantum structural model that recovers $\sigma_{\boldsymbol{A}}$ as a marginal:

$$\sigma_{\mathbf{A}} = \mathrm{Tr}_{S^{\mathrm{in}}\boldsymbol{\Lambda}}[\rho^{U}_{\mathbf{A}S|\mathbf{A}\boldsymbol{\Lambda}}(\widetilde{\tau}^{\rho_1}_{\Lambda_1} \otimes \cdots \otimes \widetilde{\tau}^{\rho_n}_{\Lambda_n})]$$

And such that $\rho^{U}_{\mathbf{A}S|\mathbf{A}\boldsymbol{\Lambda}}$ satisfies the no-influence conditions (with $\mathrm{Pa}(A_i)$ defined by G)

$$\{A_j \nrightarrow A_i\}_{A_j \notin \mathrm{Pa}(A_i)}$$



It can be shown** that:

$\sigma_{\boldsymbol{A}}$ is structurally compatible with $G \Leftrightarrow \sigma_{\boldsymbol{A}}$ is Markov for G

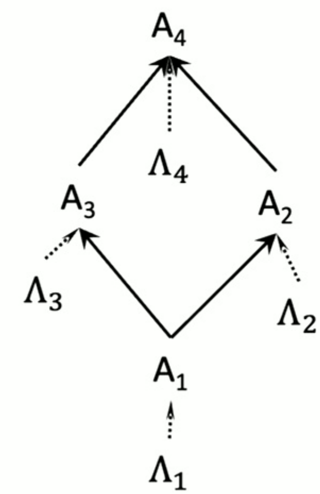
* See also slightly different definition of compatibility in Barrett *et al*, arXiv:1906.10726.

** Follows straightforwardly from proof in Barrett et al.

Quantum structural models

Given information about a particular set of outcomes $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_n)$ for the exogenous instruments, we define a *conditional process operator* *

$$\sigma_{\mathbf{A}}^{\boldsymbol{\lambda}} = \frac{\mathrm{Tr}_{S^{\mathrm{in}}\boldsymbol{\Lambda}}\left[\rho^{U}_{\mathbf{A}S|\mathbf{A}\boldsymbol{\Lambda}}(\widetilde{\tau}^{\lambda_1}_{\Lambda_1} \otimes \cdots \otimes \widetilde{\tau}^{\lambda_n}_{\Lambda_n})\right]}{P(\lambda_1, \cdots, \lambda_n)}$$



This can be used to calculate the conditional probability for a set of outcomes $\mathbf{a} = (a_1, \ldots, a_n)$ for instruments $\mathbf{z} = (z_1, \ldots, z_n)$, given $\boldsymbol{\lambda}$:

$$P_{\mathbf{z}}(\mathbf{a}|\boldsymbol{\lambda}) = \mathrm{Tr}_{\mathbf{A}}[\sigma_{\mathbf{A}}^{\boldsymbol{\lambda}} \tau_{\mathbf{A}}^{\mathbf{a}|\mathbf{z}}]$$

where we use the shorthand notation

$$\tau_{\mathbf{A}}^{\mathbf{a}|\mathbf{z}} = \tau_{A_1}^{a_1|z_1} \otimes \cdots \otimes \tau_{A_n}^{a_n|z_n}$$

* Following Costa and Shrapnel (2016).

31

Counterfactuals in QSMs

- Counterfactual Probability
  - Given a QSM M_Q, the *counterfactual probability* that outcomes $\boldsymbol{c}'$ would have obtained for a subset of nodes $\boldsymbol{C}$, had instruments $\boldsymbol{z}'$ been implemented and outcomes $\boldsymbol{b}'$ obtained at a set of nodes $\boldsymbol{B}$ (disjoint from $\boldsymbol{C}$), in the situation specified by the background variables $\boldsymbol{\Lambda} = \boldsymbol{\lambda}$, is denoted by $P^{\boldsymbol{\lambda}}_{\boldsymbol{z}'}(\boldsymbol{c}'|\boldsymbol{b}')$ and given by:

$$P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{c}'|\mathbf{b}') = \frac{P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{c}', \mathbf{b}')}{P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{b}')} = \frac{\mathrm{Tr}_{\mathbf{A}}\left[\sigma^{\lambda}_{\mathbf{A}}(\tau^{\mathbf{b}'|\mathbf{z}'_{\mathbf{B}}}_{\mathbf{B}} \otimes \tau^{\mathbf{c}'|\mathbf{z}'_{\mathbf{C}}}_{\mathbf{C}} \otimes \tau^{|\mathbf{z}'}_{\mathbf{A}\setminus\mathbf{B}\cup\mathbf{C}})\right]}{\mathrm{Tr}_{\mathbf{A}}\left[\sigma^{\lambda}_{\mathbf{A}}(\tau^{\mathbf{b}'|\mathbf{z}'_{\mathbf{B}}}_{\mathbf{B}} \otimes \tau^{|\mathbf{z}'}_{\mathbf{A}\setminus\mathbf{B}})\right]}$$

Counterfactuals in QSMs

- Counterfactual Probability
  - Given a QSM M_Q, the *counterfactual probability* that outcomes $\boldsymbol{c}'$ would have obtained for a subset of nodes $\boldsymbol{C}$, had instruments $\boldsymbol{z}'$ been implemented and outcomes $\boldsymbol{b}'$ obtained at a set of nodes $\boldsymbol{B}$ (disjoint from $\boldsymbol{C}$), in the situation specified by the background variables $\boldsymbol{\Lambda} = \boldsymbol{\lambda}$, is denoted by $P^{\boldsymbol{\lambda}}_{\boldsymbol{z}'}(\boldsymbol{c}'|\boldsymbol{b}')$ and given by:

$$P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{c}'|\mathbf{b}') = \frac{P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{c}',\mathbf{b}')}{P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{b}')} = \frac{\mathrm{Tr}_{\mathbf{A}}\left[\sigma^{\lambda}_{\mathbf{A}}(\tau^{\mathbf{b}'|\mathbf{z}'_{\mathbf{B}}}_{\mathbf{B}} \otimes \tau^{\mathbf{c}'|\mathbf{z}'_{\mathbf{C}}}_{\mathbf{C}} \otimes \tau^{|\mathbf{z}'}_{\mathbf{A}\setminus\mathbf{B}\cup\mathbf{C}})\right]}{\mathrm{Tr}_{\mathbf{A}}\left[\sigma^{\lambda}_{\mathbf{A}}(\tau^{\mathbf{b}'|\mathbf{z}'_{\mathbf{B}}}_{\mathbf{B}} \otimes \tau^{|\mathbf{z}'}_{\mathbf{A}\setminus\mathbf{B}})\right]}$$

For $P^{\lambda}_{\mathbf{z}'}(\mathbf{b}') = 0$, we set $P^{\lambda}_{\mathbf{z}'}(\mathbf{c}'|\mathbf{b}') = *$ → impossible antecedent (“counterpossible”)

32

Evaluation of counterfactuals in QCMs

- Standard quantum counterfactual query:

Given the evidence that the set of instruments $\boldsymbol{z} = (z_1, \dots, z_n)$ has been implemented and outcomes $\boldsymbol{a}$ = (a_1, …, a_n) obtained, what is the (expected) counterfactual probability $P^{\boldsymbol{a}|\boldsymbol{z}}_{\boldsymbol{b}'|\boldsymbol{z}'}(\boldsymbol{c}')$ that outcomes ***c'*** would have obtained for a subset of nodes ***C***, had instruments $\boldsymbol{z}' = (z'_1, \dots, z'_n)$ been implemented and outcomes ***b'*** obtained at a subset of nodes ***B*** (disjoint from ***C***)?

1. ***Abduction***
2. ***Action***
3. ***Prediction.***

Evaluation of counterfactuals

Given the evidence that the set of instruments $\boldsymbol{z} = (z_1, \dots, z_n)$ has been implemented and outcomes $\boldsymbol{a} = (a_1, \dots, a_n)$ obtained, what is the (expected) counterfactual probability $P^{\boldsymbol{a}|z}_{\boldsymbol{b}'|\boldsymbol{z}'}(\boldsymbol{c}')$ that outcomes $\boldsymbol{c}'$ would have obtained for a subset of nodes $\boldsymbol{C}$, had instruments $\boldsymbol{z}' = (z'_1, \dots, z'_n)$ been implemented and outcomes $\boldsymbol{b}'$ obtained at a subset of nodes $\boldsymbol{B}$ (disjoint from $\boldsymbol{C}$)?

1. Abduction

- Infer "stable events": $\{\tau^{\lambda_i}_{\Lambda_i}\}_i$

- Given observed outcomes: $\{\tau^{a_i|z_i}_{A_i}\}_i$

- Bayesian update:

$$P_{\mathbf{z}}(\boldsymbol{\lambda}|\mathbf{a}) = \frac{P_{\mathbf{z}}(\mathbf{a}|\boldsymbol{\lambda})P(\boldsymbol{\lambda})}{P_{\mathbf{z}}(\mathbf{a})} = \frac{\mathrm{Tr}_{\mathbf{A}}\left[\sigma^{\boldsymbol{\lambda}}_{\mathbf{A}}\tau^{\mathbf{a}|\mathbf{z}}_{\mathbf{A}}\right]P(\lambda_1, \cdots, \lambda_n)}{\mathrm{Tr}_{\mathbf{A}}\left[\sigma_{\mathbf{A}}\tau^{\mathbf{a}|\mathbf{z}}_{\mathbf{A}}\right]}$$

34

Evaluation of counterfactuals

Given the evidence that the set of instruments $\boldsymbol{z} = (z_1, \dots, z_n)$ has been implemented and outcomes $\boldsymbol{a} = (a_1, \dots, a_n)$ obtained, what is the (expected) counterfactual probability $P^{\boldsymbol{a}|\boldsymbol{z}}_{\boldsymbol{b}'|\boldsymbol{z}'}(\boldsymbol{c}')$ that outcomes $\boldsymbol{c}'$ would have obtained for a subset of nodes $\boldsymbol{C}$, had instruments $\boldsymbol{z}' = (z'_1, \dots, z'_n)$ been implemented and outcomes $\boldsymbol{b}'$ obtained at a subset of nodes $\boldsymbol{B}$ (disjoint from $\boldsymbol{C}$)?

2. Prediction

$$P^{\mathbf{a}|\mathbf{z}}_{\mathbf{b}'|\mathbf{z}'}(\mathbf{c}') = \sum_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} P_{\mathbf{z}}(\boldsymbol{\lambda}|\mathbf{a}) P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{c}'|\mathbf{b}')$$

Where recall that

$$P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{c}'|\mathbf{b}') = \frac{P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{c}', \mathbf{b}')}{P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{b}')} = \frac{\mathrm{Tr}_{\mathbf{A}}\left[\sigma^{\lambda}_{\mathbf{A}}(\tau_{\mathbf{B}}^{\mathbf{b}'|\mathbf{z}'_{\mathbf{B}}} \otimes \tau_{\mathbf{C}}^{\mathbf{c}'|\mathbf{z}'_{\mathbf{C}}} \otimes \tau^{|\mathbf{z}'}_{\mathbf{A}\setminus\mathbf{B}\cup\mathbf{C}})\right]}{\mathrm{Tr}_{\mathbf{A}}\left[\sigma^{\lambda}_{\mathbf{A}}(\tau_{\mathbf{B}}^{\mathbf{b}'|\mathbf{z}'_{\mathbf{B}}} \otimes \tau^{|\mathbf{z}'}_{\mathbf{A}\setminus\mathbf{B}})\right]}$$

whenever $P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{c}'|\mathbf{b}') = *$ for some $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$ with $P_{\mathbf{z}}(\boldsymbol{\lambda}|\mathbf{a}) \neq 0$, we set $P^{\mathbf{a}|\mathbf{z}}_{\mathbf{b}'|\mathbf{z}'}(\mathbf{c}') = *$ → 'counterpossible'

Evaluation of counterfactuals

Given the evidence that the set of instruments $\boldsymbol{z} = (z_1, \dots, z_n)$ has been implemented and outcomes $\boldsymbol{a} = (a_1, \dots, a_n)$ obtained, what is the (expected) counterfactual probability $P^{\boldsymbol{a}|\boldsymbol{z}}_{\boldsymbol{b}'|\boldsymbol{z}'}(\boldsymbol{c}')$ that outcomes $\boldsymbol{c}'$ would have obtained for a subset of nodes $\boldsymbol{C}$, had instruments $\boldsymbol{z}' = (z'_1, \dots, z'_n)$ been implemented and outcomes $\boldsymbol{b}'$ obtained at a subset of nodes $\boldsymbol{B}$ (disjoint from $\boldsymbol{C}$)?

2. Action

- Change instruments to $\{\tau^{a'_i|z'_i}_{A_i}\}_{a'_i}$
- Some may be do-interventions: $\tau^{\mathrm{do}(\rho)}_A \equiv (\rho_{A_{out}})^T \otimes \mathbb{I}_{A_{in}}$

Evaluation of counterfactual queries

- Classical structural model:
  - All endogenous variables determined by background variables ***u*.**

  $\Longrightarrow$ *The only way the antecedent could have been different, while keeping **u** fixed, is if some intervention had occurred.*

- Quantum structural model:
  - Complete knowledge of the exogenous variables λ does not in general determine all instrument outcomes.

  $\Longrightarrow$ *The antecedent **b'** could have occurred, even keeping all exogenous and endogenous instruments fixed, whenever:*

$$\forall \boldsymbol{\lambda} \in \boldsymbol{\Lambda} \ \left(P_{\mathbf{z}}(\boldsymbol{\lambda}|\mathbf{a}) > 0 \implies P^{\boldsymbol{\lambda}}_{\mathbf{z}'}(\mathbf{b}') > 0\right)$$

Passive vs active counterfactuals

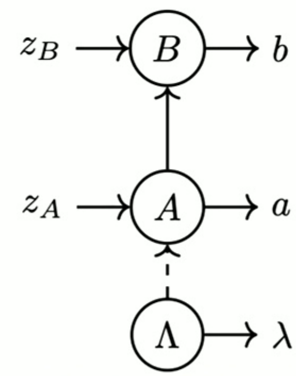
- ***Passive*** counterfactuals: no intervention is performed on the nodes specified by the antecedent;

- ***Active*** counterfactual otherwise
  - – Special case: do-interventional counterfactual

Example 1

Model M_Q: $\{\tau^\lambda\}_{\lambda=0,1} = \left\{\frac{1}{2}(([0]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}}), \frac{1}{2}(([1]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}})\right\}$

$$\rho^U_{AB|A\Lambda} = \rho^{\text{id}}_{B|A}\rho^{\text{id}}_{A|\Lambda} = \rho^{\text{id}}_{B^{\text{in}}|A^{\text{out}}}\rho^{\text{id}}_{A^{\text{in}}|\Lambda^{\text{out}}}$$



- Instruments: $\mathcal{I}_A^{z_A=1} = \{([+]_{A^{\text{out}}})^T \otimes [+]_{A^{\text{in}}}, ([-]_{A^{\text{out}}})^T \otimes [-]_{A^{\text{in}}}\}$

  $\mathcal{I}_B^{z_B} = \{\tau_B^{b|z_B}\}_b$

Q[1]: Given that *a* = + occurred with instrument $z_A = 1$, what is the probability that *b′* would have obtained for $z'_B = 1$, had it been that *a′* = - for $z'_A = z_A = 1$?

- *Passive* counterfactual query

Antecedent is possible for every value of the background variables:

$$P_{\mathbf{z}}(\lambda = 0|a = +) = \frac{1}{2} \quad \wedge \quad P^{\lambda=0}_{\mathbf{z}'}(a' = -) = \frac{1}{2},$$
$$P_{\mathbf{z}}(\lambda = 1|a = +) = \frac{1}{2} \quad \wedge \quad P^{\lambda=1}_{\mathbf{z}'}(a' = -) = \frac{1}{2}.$$

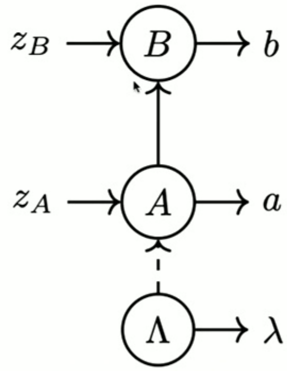
✅

40

Example 1

Model M_Q: $\{\tau^\lambda\}_{\lambda=0,1} = \left\{\frac{1}{2}(([0]_{\Lambda^{\mathrm{out}}})^T \otimes \mathbb{I}_{\Lambda^{\mathrm{in}}}), \frac{1}{2}(([1]_{\Lambda^{\mathrm{out}}})^T \otimes \mathbb{I}_{\Lambda^{\mathrm{in}}})\right\}$

$$\rho^U_{AB|A\Lambda} = \rho^{\mathrm{id}}_{B|A}\rho^{\mathrm{id}}_{A|\Lambda} = \rho^{\mathrm{id}}_{B^{\mathrm{in}}|A^{\mathrm{out}}}\rho^{\mathrm{id}}_{A^{\mathrm{in}}|\Lambda^{\mathrm{out}}}$$



- Instruments: $\mathcal{I}_A^{z_A=1} = \{([+]_{A^{\mathrm{out}}})^T \otimes [+]_{A^{\mathrm{in}}}, ([-]_{A^{\mathrm{out}}})^T \otimes [-]_{A^{\mathrm{in}}}\}$

  $\mathcal{I}_B^{z_B} = \{\tau_B^{b|z_B}\}_b$

Q$_1$: Given that *a* = + occurred with instrument $z_A = 1$, what is the probability that *b′* would have obtained for $z'_B = 1$, had it been that *a′* = - for $z'_A = z_A = 1$?

- *Passive* counterfactual query

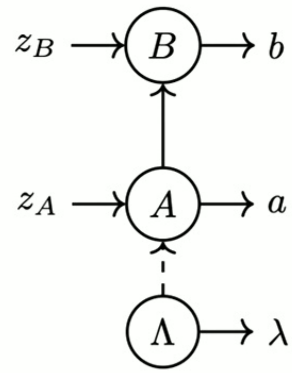
$$P^{a=+|\mathbf{z}}_{a'=-|z'_A=1}(b') = \sum_{\lambda\in\Lambda} P_{\mathbf{z}}(\lambda|a=+)P^{\lambda}_{\mathbf{z}'}(b'|a'=-)$$

$$= \mathrm{Tr}_B[[-]_{B^{\mathrm{in}}}\tau_B^{b'|z'_B=1}] \,.$$

Example 2

Model M'$_Q$: $\mathcal{I}_\Lambda^2 = \{\tau^\lambda\}_{\lambda=+,-} = \left\{\frac{1}{2}(([+]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}}), \frac{1}{2}(([-]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}})\right\}$

$\rho^U_{AB|A\Lambda} = \rho^{\text{id}}_{B|A}\rho^{\text{id}}_{A|\Lambda} = \rho^{\text{id}}_{B^{\text{in}}|A^{\text{out}}}\rho^{\text{id}}_{A^{\text{in}}|\Lambda^{\text{out}}}$



- Instruments: $\mathcal{I}_A^{z_A=1} = \{(([+]_{A^{\text{out}}})^T \otimes [+]_{A^{\text{in}}}, ([-]_{A^{\text{out}}})^T \otimes [-]_{A^{\text{in}}}\}$

  $\mathcal{I}_B^{z_B} = \{\tau_B^{b|z_B}\}_b$

- M'$_Q$ has the same process operator as M$_Q$ (they induce the same QCM), but different conditional process operators:

$$\sigma_{AB} = \rho^{\text{id}}_{B^{\text{in}}|A^{\text{out}}} \otimes \frac{1}{2}\mathbb{I}_{A^{\text{in}}}$$

$$\sigma^{\lambda=+}_{AB} = \frac{\text{Tr}_\Lambda\left[\rho^U_{AB|A\Lambda}\widetilde{\tau}^{\lambda=+}_\Lambda\right]}{P(\lambda=+)} = \rho^{\text{id}}_{B^{\text{in}}|A^{\text{out}}} \otimes [+]_{A^{\text{in}}}$$

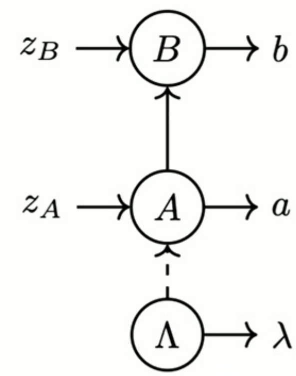
$$\sigma^{\lambda=-}_{AB} = \frac{\text{Tr}_\Lambda\left[\rho^U_{AB|A\Lambda}\widetilde{\tau}^{\lambda=-}_\Lambda\right]}{P(\lambda=-)} = \rho^{\text{id}}_{B^{\text{in}}|A^{\text{out}}} \otimes [-]_{A^{\text{in}}}$$

42

Example 2

Model M'_Q: $\mathcal{I}_\Lambda^2 = \{\tau^\lambda\}_{\lambda=+,-} = \left\{\frac{1}{2}(([+]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}}), \frac{1}{2}(([-]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}})\right\}$

$\rho^U_{AB|A\Lambda} = \rho^{\text{id}}_{B|A}\rho^{\text{id}}_{A|\Lambda} = \rho^{\text{id}}_{B^{\text{in}}|A^{\text{out}}}\rho^{\text{id}}_{A^{\text{in}}|\Lambda^{\text{out}}}$



- Instruments: $\mathcal{I}_A^{z_A=1} = \{([+]_{A^{\text{out}}})^T \otimes [+]_{A^{\text{in}}}, ([-]_{A^{\text{out}}})^T \otimes [-]_{A^{\text{in}}}\}$

  $\mathcal{I}_B^{z_B} = \{\tau_B^{b|z_B}\}_b$

Q$_1$: Given that *a* = + occurred with instrument $z_A = 1$, what is the probability that *b'* would have obtained for $z'_B = 1$, had it been that *a'* = - for $z'_A = z_A = 1$?

- Abduction:

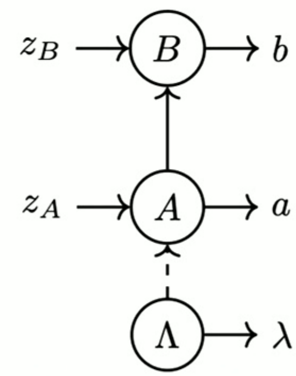
$$P_{\mathbf{z}}(\lambda = +|a = +) = 1$$
$$P_{\mathbf{z}}(\lambda = -|a = +) = 0$$

43

Example 2

Model M'_{Q}: $\mathcal{I}_{\Lambda}^{2} = \{\tau^{\lambda}\}_{\lambda=+,-} = \left\{\frac{1}{2}(([+]_{\Lambda^{\text{out}}})^{T} \otimes \mathbb{I}_{\Lambda^{\text{in}}}), \frac{1}{2}(([-]_{\Lambda^{\text{out}}})^{T} \otimes \mathbb{I}_{\Lambda^{\text{in}}})\right\}$

$\rho^{U}_{AB|A\Lambda} = \rho^{\text{id}}_{B|A}\rho^{\text{id}}_{A|\Lambda} = \rho^{\text{id}}_{B^{\text{in}}|A^{\text{out}}}\rho^{\text{id}}_{A^{\text{in}}|\Lambda^{\text{out}}}$



- Instruments: $\mathcal{I}_{A}^{z_A=1} = \{([+]_{A^{\text{out}}})^{T} \otimes [+]_{A^{\text{in}}}, ([-]_{A^{\text{out}}})^{T} \otimes [-]_{A^{\text{in}}}\}$

  $\mathcal{I}_{B}^{z_B} = \{\tau_{B}^{b|z_B}\}_{b}$

Q[1]: Given that *a* = + occurred with instrument $z_A = 1$, what is the probability that *b'* would have obtained for $z'_B = 1$, had it been that *a'* = - for $z'_A = z_A = 1$?

- Antecedent is a counterpossible for λ = +

$$P_{\mathbf{z}}(\lambda = +|a = +) = 1$$
$$P_{\mathbf{z}}(\lambda = -|a = +) = 0$$

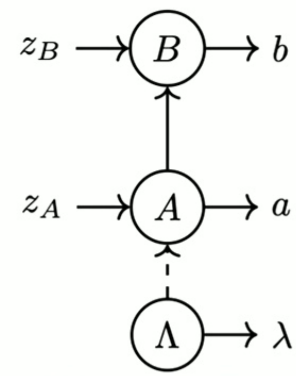
$$P^{\lambda=+}_{\mathbf{z}'}(b'|a' = -) = \frac{P^{\lambda=+}_{\mathbf{z}'}(b', a' = -)}{P^{\lambda=+}_{\mathbf{z}'}(a' = -)} = *$$

44

Example 2

Model M'_Q: $\mathcal{I}^2_\Lambda = \{\tau^\lambda\}_{\lambda=+,-} = \left\{\frac{1}{2}(([+]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}}), \frac{1}{2}(([-]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}})\right\}$

$$\rho^U_{AB|A\Lambda} = \rho^{\text{id}}_{B|A}\rho^{\text{id}}_{A|\Lambda} = \rho^{\text{id}}_{B^{\text{in}}|A^{\text{out}}}\rho^{\text{id}}_{A^{\text{in}}|\Lambda^{\text{out}}}$$



- Instruments: $\mathcal{I}^{z_A=1}_A = \{([+]_{A^{\text{out}}})^T \otimes [+]_{A^{\text{in}}}, ([-]_{A^{\text{out}}})^T \otimes [-]_{A^{\text{in}}}\}$

  $\mathcal{I}^{z_B}_B = \{\tau^{b|z_B}_B\}_b$

Q_1: Given that *a* = + occurred with instrument $z_A = 1$, what is the probability that *b'* would have obtained for $z'_B = 1$, had it been that *a'* = - for $z'_A = z_A = 1$?

- → Expected CF probability is also not well-defined

$$P_{\mathbf{z}}(\lambda = +|a = +) = 1$$
$$P_{\mathbf{z}}(\lambda = -|a = +) = 0$$

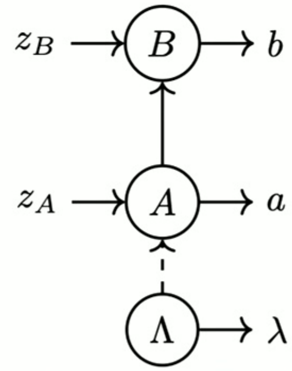
$$P^{\lambda=+}_{\mathbf{z}'}(b'|a' = -) = \frac{P^{\lambda=+}_{\mathbf{z}'}(b', a' = -)}{P^{\lambda=+}_{\mathbf{z}'}(a' = -)} = *$$

$$\implies P^{a=+|\mathbf{z}}_{a'=-|z'_A=1}(b') = \sum_{\lambda\epsilon\Lambda} P_{\mathbf{z}}(\lambda|a = +)P^{\lambda}_{\mathbf{z}'}(b'|a' = -) = *$$

Example 2

Model M'$_Q$: $\mathcal{I}^2_\Lambda = \{\tau^\lambda\}_{\lambda=+,-} = \left\{\frac{1}{2}(([+]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}}), \frac{1}{2}(([-]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}})\right\}$

$$\rho^U_{AB|A\Lambda} = \rho^{\text{id}}_{B|A}\rho^{\text{id}}_{A|\Lambda} = \rho^{\text{id}}_{B^{\text{in}}|A^{\text{out}}}\rho^{\text{id}}_{A^{\text{in}}|\Lambda^{\text{out}}}$$



- Instruments: $\mathcal{I}^{z_A=1}_A = \{([+]_{A^{\text{out}}})^T \otimes [+]_{A^{\text{in}}}, ([-]_{A^{\text{out}}})^T \otimes [-]_{A^{\text{in}}}\}$

  $\mathcal{I}^{z_B}_B = \{\tau^{b|z_B}_B\}_b$ $\quad$ $\boxed{\mathcal{I}^{z'_A=2}_A = \tau^{\text{do}([-])}_A = ([-]_{A^{\text{out}}})^T \otimes \mathbb{I}_{A^{\text{in}}}}$

Q$_1$: Given that *a* = + occurred with instrument $z_A = 1$, what is the probability that *b'* would have obtained for $z'_B = 1$, had it been that *a'* = - for the do-instrument $z_A = 2$?

- Abduction probabilities are the same, and CF probabilities are well-defined:

$$P_{\mathbf{z}}(\lambda = +|a = +) = 1$$
$$P_{\mathbf{z}}(\lambda = -|a = +) = 0$$

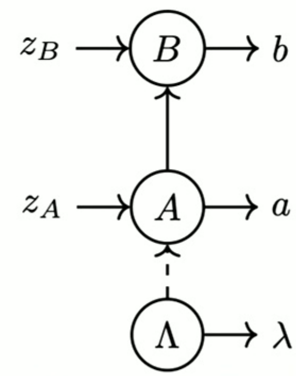
$$P^\lambda_{\mathbf{z}'}(b'|a' = -) = \frac{P^\lambda_{\mathbf{z}'}(b', a' = -)}{P^\lambda_{\mathbf{z}'}(a' = -)} = P^\lambda_{\mathbf{z}'}(b', a' = -)$$

46

Example 2

Model M'_{Q}: $\mathcal{I}^2_\Lambda = \{\tau^\lambda\}_{\lambda=+,-} = \left\{\frac{1}{2}(([+]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}}), \frac{1}{2}(([-]_{\Lambda^{\text{out}}})^T \otimes \mathbb{I}_{\Lambda^{\text{in}}})\right\}$

$$\rho^U_{AB|A\Lambda} = \rho^{\text{id}}_{B|A}\rho^{\text{id}}_{A|\Lambda} = \rho^{\text{id}}_{B^{\text{in}}|A^{\text{out}}}\rho^{\text{id}}_{A^{\text{in}}|\Lambda^{\text{out}}}$$



- Instruments: $\mathcal{I}^{z_A=1}_A = \{([+]_{A^{\text{out}}})^T \otimes [+]_{A^{\text{in}}}, ([-]_{A^{\text{out}}})^T \otimes [-]_{A^{\text{in}}}\}$

  $\mathcal{I}^{z_B}_B = \{\tau^{b|z_B}_B\}_b$ $\quad$ $\boxed{\mathcal{I}^{z'_A=2}_A = \tau^{\text{do}([-])}_A = ([-]_{A^{\text{out}}})^T \otimes \mathbb{I}_{A^{\text{in}}}}$

Q_1: Given that *a* = + occurred with instrument $z_A = 1$, what is the probability that *b'* would have obtained for $z'_B = 1$, had it been that *a'* = - for the do-instrument $z_A = 2$?

- Expected CF probability can be computed as

$$P_{\mathbf{z}}(\lambda = +|a = +) = 1$$
$$P_{\mathbf{z}}(\lambda = -|a = +) = 0$$

$$P^\lambda_{\mathbf{z}'}(b'|a' = -) = \frac{P^\lambda_{\mathbf{z}'}(b', a' = -)}{P^\lambda_{\mathbf{z}'}(a' = -)} = P^\lambda_{\mathbf{z}'}(b', a' = -)$$

$$\Longrightarrow \quad P^{a=+|\mathbf{z}}_{a'=-|z'_A=2}(b') = P^{\lambda=+}_{\mathbf{z}'}(b'|a' = -) = \text{Tr}_B[[-]_{B^{\text{in}}}\tau^{b'|z'_B=2}_B]$$

47

Passive vs Active?

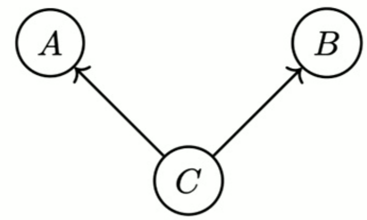
- Why not just only consider do-interventional CF queries?
  - Lewis, Pearl: "minimal modification", "closest possible world"
  - *No* modification to the model is, by definition, the *minimal* modification.
  - Counterfactual antecedent as a do-intervention is a *different event* (different CP map)

- *Principle of Minimality:*

  If it is ambiguous whether a CF query is intended as a passive or active CF, it should be interpreted passively if it is not a counterpossible, that is, if its antecedent is not impossible.

Example: Bell scenario

$$\mathcal{I}_A = \{([0]_{A^{\rm out}})^T \otimes [0]_{A^{\rm in}}, ([1]_{A^{\rm out}})^T \otimes [1]_{A^{\rm in}}\}$$
$$\mathcal{I}_B = \{([0]_{B^{\rm out}})^T \otimes [0]_{B^{\rm in}}, ([1]_{B^{\rm out}})^T \otimes [1]_{B^{\rm in}}\}$$
$$\mathcal{I}_C = \left\{([\Phi_+]_{C^{\rm out}})^T \otimes \mathbb{I}_{C^{\rm in}}\right\}, \ |\Phi_+\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle_{C_A^{\rm out}}|0\rangle_{C_B^{\rm out}} + |1\rangle)_{C_A^{\rm out}}|1\rangle_{C_B^{\rm out}}$$
$$\rho^U_{AB|C} = \rho^{\rm id}_{A|C_A^{\rm out}}\rho^{\rm id}_{B|C_B^{\rm out}}$$



Q1: Given that *a =b= 0*, what's the probability that *b' = 1* had it been that *a' = 1*?

Q2: Given that *a =b= 0*, what's the probability that *b' = 1* had it been that *a' = 0*?

1) Interpret as a do-interventional CFs:

$$P^{a=b=0}_{{\rm do}(a'=0)}(b'=1) = P_{{\rm do}(a'=1)}(b'=1|a'=1) = \frac{1}{2}$$
$$P^{a=b=0}_{{\rm do}(a'=1)}(b'=1) = P_{{\rm do}(a'=1)}(b'=1|a'=1) = \frac{1}{2}$$

2) Interpret as a passive CF:

$$P^{a=b=0|\mathbf{z}}_{a'=1|\mathbf{z}'=\mathbf{z}}(b'=1) = P_{\mathbf{z}'=\mathbf{z}}(b'=1|a'=1) = 1$$
$$P^{a=b=0|\mathbf{z}}_{a'=0|\mathbf{z}'=\mathbf{z}}(b'=1) = P_{\mathbf{z}'=\mathbf{z}}(b'=1|a'=0) = 0$$

→ In QCMs there can be counterfactual dependence without causal dependence

49

Counterfactual Definiteness

- QCMs violate "counterfactual definiteness" → is this the lesson of Bell's theorem?
  - Determinism is not *required* for deriving a Bell inequality
  - can occur in "merely" indeterministic (but otherwise classical) models
- CF dependence without causal dependence, we suggest, better captures a nonclassical feature of QCMs

Counterfactual Definiteness

- QCMs violate "counterfactual definiteness" → is this the lesson of Bell's theorem?
  - Determinism is not *required* for deriving a Bell inequality
  - can occur in "merely" indeterministic (but otherwise classical) models
- CF dependence without causal dependence, we suggest, better captures a nonclassical feature of QCMs

Thank you!