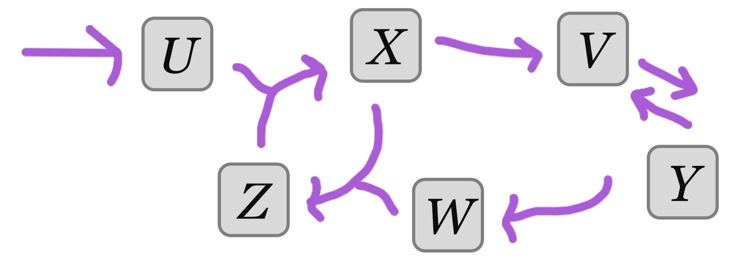
Qualitative Mechanism Independence

Oliver E. Richardson,
Spencer Peters,
Joseph Y. Halpern

Cornell Bowers CIS Computer Science

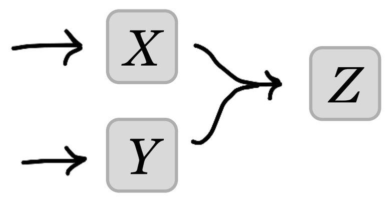
NEURAL INFORMATION PROCESSING SYSTEMS

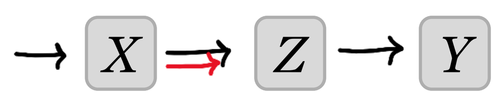
Describe qualitative aspects of probability like (conditional) (in)dependence with (directed) hypergraphs, whose arcs represent *independent mechanisms*.



Examples.

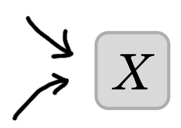
X and Y are independent



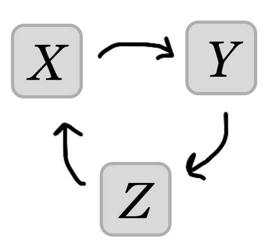


X and Y are conditionally independent given Z
(and Z is a function of X)

X is constant
(i.e., X and X are independent)



Q: What should



mean ?

$\mu \models \Diamond \mathcal{A}$

Definition. A joint distribution $\mu(\mathcal{X})$ is (QIM-)*compatible* with a hypergraph $\mathcal{A}$ iff:

$\exists$ extension ("witness") $\bar{\mu}(\mathcal{X}, \mathcal{U})$ of $\mu(\mathcal{X})$ (a)

to mutually independent "noise" variables $\{ U_a \}_{a \in \mathcal{A}}$ (b)

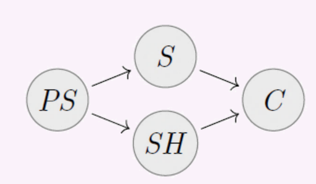
such that, $\forall$ arcs $S \xrightarrow{a} T \in \mathcal{A}$, (c)

U_a and source vars $S_{\subseteq \mathcal{X}}$ determine target vars $T_{\subseteq \mathcal{X}}$.

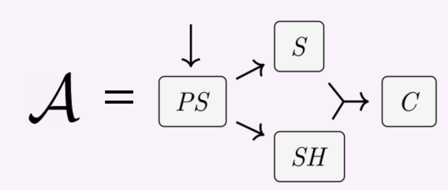
Some Nice Properties:

1. Generalizes independencies of Bayesian Networks

e.g., μ has the independences of the BN



if and only if μ is compatible with



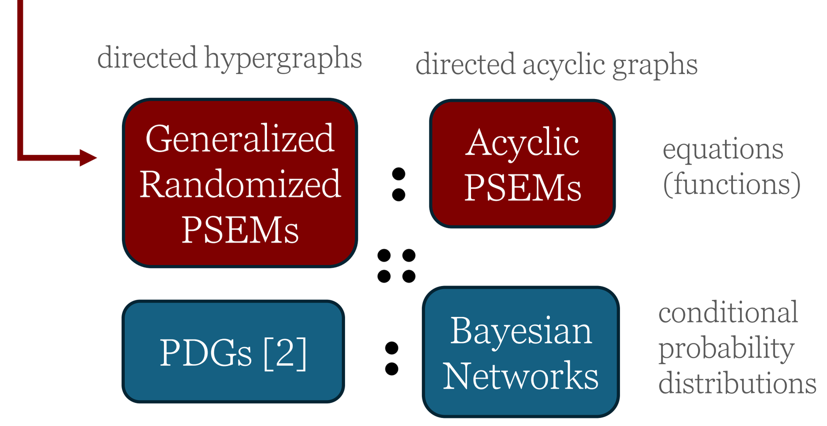
2. Captures arbitrary functional dependencies with parallel arcs
3. Gives meaning to cyclic and over-constrained models...

Causality

Establishing a causal interpretation of QIM-compatibility, by relating it directly to causal models.

❖ $\mu \models \Diamond \mathcal{A}$ $\iff$ μ can arise in a randomized causal model with dependency structure $\mathcal{A}$.

For $\mathcal{A}$ with overlapping targets, need a generalization of SEMs, related to *causal models with constraints [1]*.



❖ While the above says their existence coincides, witnesses $\bar{\mu}$ & causal models M are quite different:

$\bar{\mu}$ doesn't say anything about probability-zero counterfactuals;

meanwhile, M doesn't specify choice between fixed points.

Yet the two fit together in an elegant (adjunctive) way:

M can be derived from $\bar{\mu}$ $\iff$ $\bar{\mu}$ satisfies eqns of M

❖ The relationship extends to interventions. Perhaps surprisingly, when M and $\bar{\mu}$ have the relationship described above,

intervening in M corresponds to conditioning on an event in $\bar{\mu}$!

Concretely, there is an event $do_M(\mathbf{X}{=}\mathbf{x})$ involving noise variables, and conditioning on this event has the effect of intervening in M.

$$\Pr_{\mathcal{M}}([\mathbf{X} \leftarrow \mathbf{x}]\varphi) \le \bar{\mu}(\varphi \,|\, \mathrm{do}_{\mathcal{M}}(\mathbf{X}{=}\mathbf{x})) \le \Pr_{\mathcal{M}}(\langle \mathbf{X} \leftarrow \mathbf{x} \rangle \varphi)$$

"after the intervention $\mathbf{X}{=}\mathbf{x}$, φ is true in all contexts"

"after the intervention $\mathbf{X}{=}\mathbf{x}$, φ is true in some context"

Measuring μ's *degree of QIM-incompatibility:*

$$\inf_{\substack{\nu(\mathcal{U},\mathcal{X})\\ \nu(\mathcal{X})=\mu(\mathcal{X})}} \; -\mathrm{H}_\nu(\mathcal{U}) + \sum_{a \in \mathcal{A}} \mathrm{H}_\nu(U_a) + \sum_{a \in \mathcal{A}} \mathrm{H}_\nu(T_a | S_a, U_a) \quad =: \mathrm{QIM}Inc_{\mathcal{A}}(\mu)$$

(a) (b) (c)

≥ 0 with equality iff $\mu \models \Diamond \mathcal{A}$

Conclusions and Open Questions

The beginnings of a bridge between causality and information theory, based on complex dependency structures. Payoffs:

- A simple information-theoretic test for cyclic causal structures;
- Clarifying causal picture of information-theoretic primitives;
- New justification of the qualitative PDG scoring function, IDef.

Many basic questions remain open.

- *Are the probabilities consistent with clockwise and counter-clockwise cycles?*
- *What is the complexity of deciding whether or not* $\mu \models \Diamond \mathcal{A}$ *?*

Information Theory

Beyond independence and dependence, QIM-compatibility can express generalizations of both notions, related in information theory.

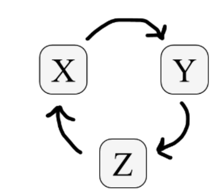
Information Deficiency, $IDef_{\mathcal{A}}(\mu)$ — the qualitative scoring function for Probabilistic Dependency Graphs (PDGs) [2]

Theorem.

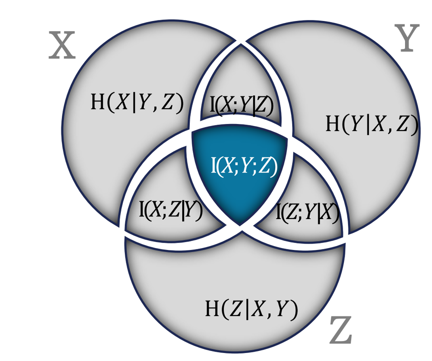
$$\mu \models \Diamond \mathcal{A} \implies -\mathrm{H}_\mu(\mathcal{X}) + \sum_{a \in \mathcal{A}} \mathrm{H}_\mu(T_a \mid S_a) \le 0$$

A deep information-theoretic consequence of arbitrary structures;

1. Generalizes independence property of BNs
2. Captures functional dependence properties
3. Application to the 3-cycle



resolves longstanding confusion about *interaction information:*

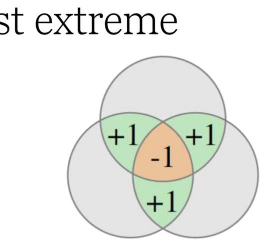


Contrary to accepted wisdom [3],

I(X;Y;Z) < 0 implies a *(causal) 3-way interaction,*

validating "novice" intuitions from the most extreme example: *the xor distribution*

in which X and Y are independent fair coins, and Z = X ⊕ Y .



Generalization (*IDef* Bounds).

For all extensions $\bar{\mu}(\mathcal{X}, \mathcal{U})$ of $\mu(\mathcal{X})$

$$IDef_{\mathcal{A}}(\mu) \le \mathrm{QIM}Inc_{\mathcal{A}}(\mu) \le IDef_{\mathcal{A}^\dagger}(\bar{\mu})$$

also, there is some $\bar{\mu}$ for which this is an equality

the "augmented" hypergraph obtained by explicitly modeling noise, i.e.,

replacing each arc $S \xrightarrow{a} T$ with U_a, $S \rightarrow T$,

and also including a "global" mechanism $\mathcal{U} \rightarrow \mathcal{X}$.

"Flipping" a standard counter-example:

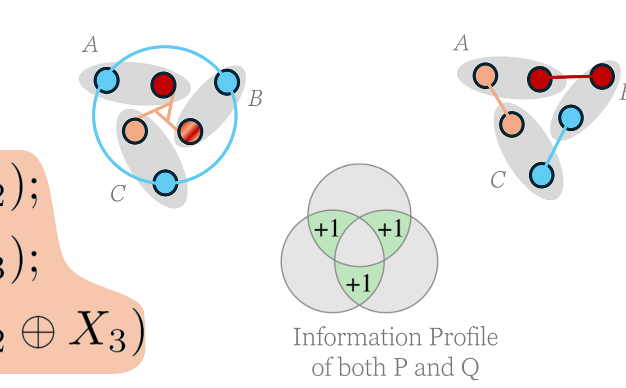
two distributions with the same information profile, but, allegedly, fundamentally different structural properties.

$P(A,B,C)$

$A := (X_1, X_2);$
$B := (X_1, X_3);$
$C := (X_1, X_2 \oplus X_3)$

seemingly irrelevant shared bit

xor distribution; has irreducible "3-way interaction"

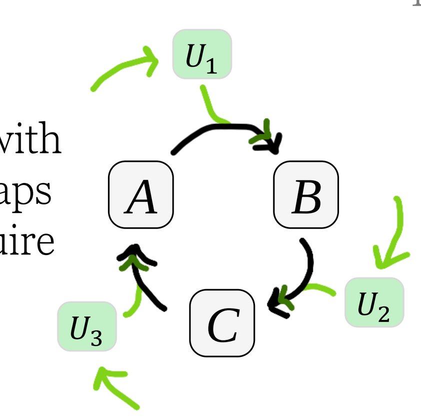


$Q(A,B,C)$

$A := (X_1, X_2);$
$B := (X_2, X_3);$
$C := (X_3, X_1);$

Only "2-way" Interactions

Yet P and Q are compatible with the same hypergraphs! Perhaps surprisingly, P does not require joint causal dependence:



$U_1 = X_3 \oplus X_1,$
$U_2 = X_2,$ and
$U_3 = X_3$

References.

[1] S. Beckers, J.Y.Halpern, and C.Hitchcock. *Causal Models with Constraints*, 2023
[2] O.E.Richardson and J. Y. Halpern, *Probabilistic Dependency Graphs*. In Proceedings: AAAI 2021, pages 12174-12181.
[3] R. G. James and J. P. Crutchfield. *Multivariate Dependence Beyond Shannon Information*. Entropy, 19(10), 2017. ISSN 1099-4300.