Causal models in string diagrams (Extended abstract)

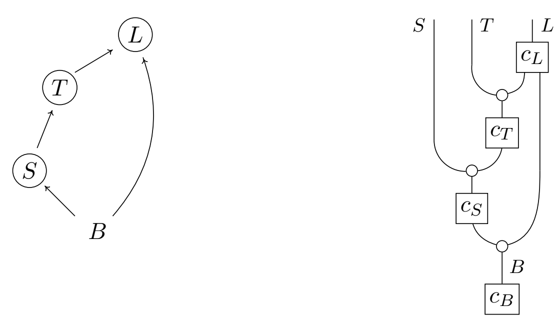
Robin Lorenz, Sean Tull
Quantinuum, 17 Beaumont Street, Oxford, UK

This is a summary of findings from recent work, available as a preprint linked in Ref. [1], that presents many aspects of causal reasoning according to the causal model framework in a string diagrammatic language, based on a generalised category- theoretic notion of a causal model.

The framework of *causal models*, pioneered by Pearl [2], and Spirtes, Glymour and Scheines [3], provides a principled approach to causal reasoning, applied today across many scientific domains. Given a set of variables V, the framework spells out how causal structure, typically represented by a *directed acyclic graph* (DAG), and empirical data over V (either observational or interventional in nature) constrain each other. These stipulations are encoded in the notion of a causal model, which underlies the framework's achievements such as *causal discovery* algorithms and solutions to *causal inference* problems. These include solving, given causal knowledge but incomplete data, when and how can one predict what would happen if one *intervened* on the variables, or when one can unambigously answer *counterfactual* questions.

Despite its successes, the conventional presentation of causal models can be unwieldy in that it is formulated in terms of both DAGs (or more general graphs) and probability distributions. While the properties of one may licence certain calculations in terms of the other, in a sense these stay separate, leading to fairly involved bookkeeping to keep track of how both objects constrain each other.

Building on *categorical probability theory* [4, 5, 6, 7] based on working in a *symmetric monoidal category* with a 'copy-discard' structure (*cd-category*), and using the fact that DAGs correspond precisely to a certain class of string diagrams in a cd-category, Refs. [8, 9] showed how a causal model can be presented entirely in terms of cd-categories (see [1] for a more detailed overview on prior works). Essentially, a causal model with causal structure as on the left below (encircled vertices correspond to 'observed variables') can be represented by a string diagram as on the right, which we call a *network diagram*, with the boxes interpreted in a cd-category (say, finite sets and stochastic maps). These morphisms, the *causal mechanisms*, become the core data of a model.



(1)

Based on this correspondence, there is a growing movement in applying categorical methods to causal reasoning, including [9, 10, 11]. However, so far most works have only focused on specific aspects of causality and may not be accessible to typical causal inference researchers. As such a more systematic account of the framework would be desirable.

This work [1] builds on these developments to present a more comprehensive and accessible string diagrammatic treatment of the causal model framework as a whole, which we hope to be useful to researchers in both causality and applied category theory. In particular, this work:

- presents diagrammatic definitions of causal models (as in (1)) and *functional* causal models in a cd-category, generalising *causal Bayesian networks* and *structural causal models*, respectively.
- treats *interventions* on a model in full generality, beyond simply atomic *do-interventions*, including soft and conditional interventions which are naturally represented in diagrams.

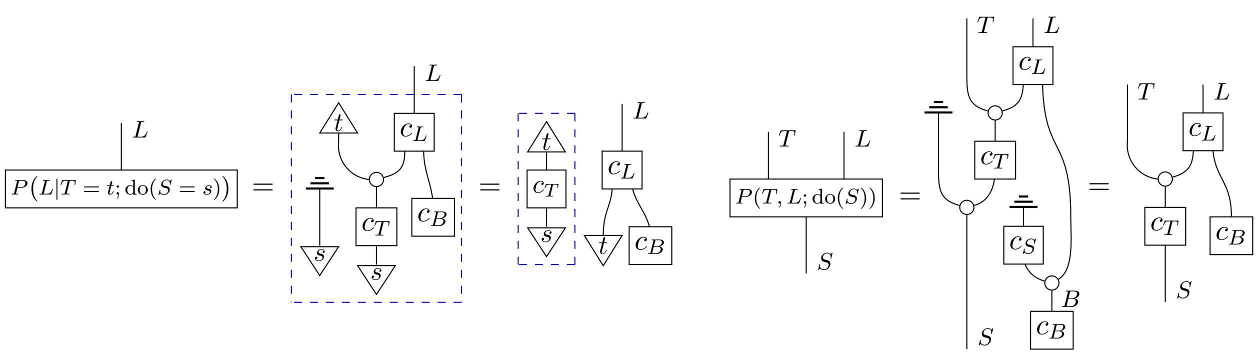
- introduces a novel diagrammatic formalism for *Bayesian conditioning*, allowing causal inference calculations to be done entirely diagrammatically. This is based on a *normalisation box*, satisfying graphical axioms, which turns any morphism into a *partial channel* and makes use of working in a general cd-category, rather than only its Markov category of channels as in prior approaches.

- introduces the notion of *open causal model*, or causal model 'with inputs' [1] – together with their transformations, yielding a category of (open) causal models within any cd-category. Open models naturally arise from certain interventions, and describe the compositional 'building blocks' for typical (closed) causal models.

- discusses how to go from a latent variable model based on an *acyclic directed mixed graph*, which captures latent common causes, to a string diagrammatic representation. This is a key practical step in causal inference calculations which has not yet been discussed in categorical approaches.

- discusses *causal effect identifiability*, mildly extending the work in Ref. [9] and incorporating our diagrammatic treatment of conditioning, which reveals subtleties between concluding identifiability and assumptions of full support that seem to be glossed over in the conventional literature.

- treats *counterfactuals* in a general cd-category. This includes a diagrammatic definition of counterfactuals which is of clarificatory value, as well as a diagrammatic treatment of the problem of the *identifiability of counterfactuals*, including analogues of the well-known `make-cg` and `IDC*` algorithms, which we argue to be more accessible in string-diagrammatic terms.



(a) (b)

Figure 1: Examples using the above model; (a) do-intervention $\mathrm{do}(S = s)$ with conditioning on $T = t$ using the normalisation box; (b) the open model arising from $\mathrm{do}(S)$.

In summary, this work aims to bring the treatments of probability theory and causality developed in the applied category theory community to a wider audience of causal model practitioners, by demonstrating its use in a variety of aspects of the framework including interventions, counterfactuals and identifability problems across all levels of 'Pearl's hierarchy' [12]. Overall, it argues and demonstrates that *causal reasoning* according to the causal model framework is most naturally and intuitively done as *diagrammatic reasoning*.

References

[1] R. Lorenz and S. Tull, "Causal models in string diagrams," *arXiv preprint arXiv:2304.07638* (2023) . `https://arxiv.org/abs/2304.07638`.

[2] J. Pearl, *Causality*. Cambridge university press, 2009.

[3] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*. MIT press, 2nd ed., 2000.

[1] These are related to the 'generalised causal models' of [7] but in particular only allow single output boxes, to allow for a causal interpretation in the usual sense.

[4] B. Jacobs and F. Zanasi, "The logical essentials of bayesian reasoning," *Foundations of Probabilistic Programming* (2019) 295–331.

[5] K. Cho and B. Jacobs, "Disintegration and bayesian inversion via string diagrams," *Mathematical Structures in Computer Science* **29** no. 7, (2019) 938–971.

[6] T. Fritz, "A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics," *Advances in Mathematics* **370** (2020) 107239.

[7] T. Fritz and A. Klingler, "The d-separation criterion in categorical probability," *Journal of Machine Learning Research* **24** no. 46, (2023) 1–49.

[8] B. Fong, "Causal theories: A categorical perspective on bayesian networks," *arXiv preprint arXiv:1301.6201* (2013) .

[9] B. Jacobs, A. Kissinger, and F. Zanasi, "Causal inference via string diagram surgery: A diagrammatic approach to interventions and counterfactuals," *Mathematical Structures in Computer Science* **31** no. 5, (2021) 553–574.

[10] T. Fritz and W. Liang, "Free gs-monoidal categories and free markov categories," *arXiv preprint arXiv:2204.02284* (2022) .

[11] T. Fritz and A. Klingler, "The d-separation criterion in categorical probability," *arXiv preprint arXiv:2207.05740* (2022) .

[12] E. Bareinboim, J. D. Correa, D. Ibeling, and T. F. Icard, "On Pearl's hierarchy and the foundations of causal inference," 2021.