Causal Spaces and String Diagrams

Siddharth Setlur

November 3, 2025

1 Introduction

We aim to clarify the definitions and results in [Par+24] and place causal spaces within existing category theoretic frameworks for causality.

2 Causal Spaces

Definition 2.1 (Causal Space). *A causal space is a product probability space* $(\Omega, \mathcal{H}, \mathbb{P}) = (\times_{t \in T}\Omega_t, \otimes_{t \in T}\mathcal{E}_t, \mathbb{P})$ *equipped with a collection of Markov kernels* $\mathbb{K} = \{K_S : S \in 2^T\}$ *called a* causal mechanism. *The Markov kernels* $K_S : (\Omega_S, \mathcal{H}_S) \to (\Omega, \mathcal{H})$ *satisfy the following two axioms:*

1. *for all* $A \in \mathcal{H}$ *and* $\omega_s \in \Omega_s$,
$$K_\emptyset(\omega_S, A) = \mathbb{P}(A);$$
2. *for all* $\omega_S \in \Omega_S$, $A \in \mathcal{H}_S$, *and* $B \in \mathcal{H}$,
$$K_S(\omega_S, A \cap B) = 1_A(\omega_S)K_S(\omega_S, B) = \delta_{\omega_S}(A)K_S(\omega_S, B).$$
In particular, for $A \in \mathcal{H}_S$, $K_S(\omega_s, A) = \delta_{\omega_S}(A)$.

Remark 2.1 (Observational and interventional distributions). In Definition 2.1, the probability measure $\mathbb{P}$ plays the role of the *observational distribution* in the SCM framework. The causal mechanisms provide a means of computing interventional distributions. In particular, the kernel K_S allows us to determine what happens when we intervene on a subset $S \in 2^T$ of variables of the system.

Viewed through this lens, axiom 1 simply says that intervening on the empty set (not intervening at all) has no effect (the distribution on any event A after not intervening is just the observational distribution $\mathbb{P}(A)$). Axiom 2 is a bit trickier to unravel, so we rephrase it below using the intuition developed in a simple example we consider next.

Example 1 (Two variable binary causal space). Let $T = \{X, Y\}$ be a set of two variables and $\Omega_X = \Omega_Y = \{0, 1\}$. Consider the product probability space $(\Omega_X \times \Omega_Y, \mathcal{E}_X \otimes \mathcal{E}_Y, \mathbb{P})$, where $\mathcal{E}_X$ and $\mathcal{E}_Y$ are the power sets of Ω_X and Ω_Y, respectively and let $\mathbb{P}$ be the uniform distribution on Ω. Axiom 2 constrains K_X and K_Y. Recall that a Markov kernel K between discrete spaces $(\Omega, \mathcal{H})$ and $(\Omega', \mathcal{H}')$ can be written as a $|\Omega'| \times |\Omega\|$ matrix, $K_{ij} = K(\omega'_i, \omega_j)$. Thus, we can write

$$K_X = \begin{pmatrix} K_X(0,(0,0)) & K_X(1,(0,0)) \\ K_X(0,(0,1)) & K_X(1,(0,1)) \\ K_X(0,(1,0)) & K_X(1,(1,0)) \\ K_X(0,(1,1)) & K_X(1,(1,1)). \end{pmatrix}$$

Axiom 2 forces K_X to be a block matrix of the form

$$K_X = \left(\begin{array}{c|c} \begin{matrix} K_X(0,(0,0)) \\ K_X(0,(0,1)) \end{matrix} & \mathbf{0} \\ \hline \mathbf{0} & \begin{matrix} K_X(1,(1,0)) \\ K_X(1,(1,1)) \end{matrix} \end{array}\right)$$

Indeed, we have $A = \{(0,0),(0,1)\} \in \mathcal{H}_X$ and so (by axiom 2)

$$K_X(\omega_x, A) = \delta_{\omega_X}(A) = \begin{cases} 1 & \text{if } \omega_X = 0, \\ 0 & \text{if } \omega_X = 1. \end{cases}$$

Similarly, for $B = \{(1,0),(1,1)\} \in \mathcal{H}_X$, we have

$$K_X(\omega_x, B) = \delta_{\omega_X}(B) = \begin{cases} 1 & \text{if } \omega_X = 1, \\ 0 & \text{if } \omega_X = 0. \end{cases}$$

This yields the block structure above since $K_X(0,B) = 0$ and $K_X(1,A) = 0$ while the other two entries are non-zero and must sum to 1. Intuitively, $K_S(0,A)$ is the probability distribution on A after intervening by setting $X = 0$. If A does not contain an event where $X = 0$, then the intervention makes A impossible, hence the probability is 0. Conversely, if A contains all the events where $X = 0$, then after intervening to set $X = 0$, the probability of A is 1.

Lemma 2.1 (Rephrasing axiom 2). *Axiom 2 of Definition 2.1 can be rephrased as follows: for all $\omega_S \in \Omega_S$ and $B \in \mathcal{H}$:*

$$K_S(\omega_S, B) = \begin{cases} 0 & \text{if } \omega_S \notin \pi_S(B) \\ K_S(\omega_S, B \cap \{\Omega_{T\setminus S} \times \omega_S\}) & \text{if } \omega_S \in \pi_S(B). \end{cases}$$

where $\pi_S(B)$ denotes the projection of B onto Ω_S Sid: I think we only need case 2. If $\omega_S \notin \pi_S(B)$, then $B \cap \{\Omega_{T\setminus S} \times \omega_S\} = \emptyset$ and so $K_S(\omega_S, B \cap \{\Omega_{T\setminus S} \times \omega_S\}) = 0$ anyway.

Proof. We first show that axiom 2 implies the statement above. Let $\omega_S \in \Omega_S$ and $B \in \mathcal{H}$ be arbitrary. Suppose $\omega_S \notin \pi_S(B)$ and set $A = \{\omega_S\} \times \Omega_{T\setminus S} \in \mathcal{H}_S$. Applying axiom 2, we get

$$K_S(\omega_S, A \cap B) = \delta_{\omega_S}(A) K_S(\omega_S, B).$$

Due to $\omega_S \notin \pi_S(B)$, we have $A \cap B = \emptyset$ and consequently $K_S(\omega_S, A \cap B) = 0$. On the right hand side, we have $\delta_{\omega_S}(A) = 1$ since $\omega_S \in A$. Thus, we conclude that $K_S(\omega_S, B) = 0$.

Suppose now that $\omega_S \in \pi_S(B)$ and set $A = \{\omega_S\} \times \Omega_{T\setminus S} \in \mathcal{H}_S$. Applying axiom 2 again, we get

$$K_S(\omega_S, A \cap B) = \delta_{\omega_S}(A) K_S(\omega_S, B) = K_S(\omega_S, B).$$

Conversely, suppose the statement above holds and let $\omega_S \in \Omega_S$, $A \in \mathcal{H}_S$, and $B \in \mathcal{H}$ be arbitrary. We have two cases to consider. First, consider the case where $\omega_S \in A$. We need to use the rephrased axiom to show

$$K_S(\omega_S, A \cap B) = K_S(\omega_S, B)$$

(since $\delta_{\omega_S}(A) = 1$). If $A \cap B = \emptyset$, then in particular $\omega_S \notin \pi_S(B)$ (since $\omega_S \in A$ by assumption). Thus, by the rephrased axiom, we have $K_S(\omega_S, B) = 0 = K_S(\omega_S, \emptyset)$. Assume $A \cap B \neq$ and apply the rephrased axiom to this set noting that by construction $\omega_S \in \pi_S(A \cap B)$.

$$\begin{aligned} K_S(\omega_S, A \cap B) &= K_S(\omega_S, B \cap A \cap \{\omega_S \times \Omega_{T \setminus S}\}) \\ &\overset{\spadesuit}{=} K_S(\omega_S, B \cap \{\omega_S \times \Omega_{T \setminus S}\}) \\ &\overset{\diamondsuit}{=} K_S(\omega_S, B). \end{aligned}$$

where ($\spadesuit$) follows from the fact that $A \cap \{\omega_S \times \Omega_{T\setminus S}\} = \{\omega_S \times \Omega_{T \setminus S}\}$ (since $\omega_S \in A$) and ($\diamondsuit$) follows from applying the rephrased axiom.Sid: The final step is more elgant if we use the definition of the axiom where we don't have the two cases (see my comment above). If we do have the two cases we just need to make a comment in the $\diamondsuit$ step that if the intersection is empty we get 0 but this is OK since the interesection being empty means that $\omega_S \notin \pi_S(B)$ and so we get $K_S(\omega, B)$ eitherway. On the other hand, if $\omega_S \notin A$, we need to show that

$$K_S(\omega_S, A \cap B) = 0.$$

Let $\tilde{B} = A \cap B$. We can assume without loss of generality that $\tilde{B} = \emptyset$, since there is nothing to show otherwrise (by definition we have $K_S(\omega_S, A \cap B) = 0$). Note that $\omega_S \notin \pi(\tilde{B})$, since $\omega_S \notin A$. So, applying the rephrased axiom to $\tilde{B}$, we get $K_S(\omega_S, \tilde{B}) = 0$ as desired. ■

References

[BPS24] Simon Buchholz, Junhyung Park, and Bernhard Schölkopf. *Products, Abstractions and Inclusions of Causal Spaces*. 2024. DOI: 10.48550/ARXIV.2406.00388. URL: https://arxiv.org/abs/2406.00388 (visited on 05/10/2025).

[FK23] Tobias Fritz and Andreas Klingler. "The d-Separation Criterion in Categorical Probability". In: *Journal of Machine Learning Research* 24.46 (2023), pp. 1–49. ISSN: 1533-7928. URL: http://jmlr.org/papers/v24/22-0916.html (visited on 08/01/2025).

[Gei+21] Atticus Geiger et al. *Causal Abstractions of Neural Networks*. 2021. DOI: 10.48550/ARXIV.2106.02997. URL: https://arxiv.org/abs/2106.02997.

[Gei+25] Atticus Geiger et al. "Causal Abstraction: A Theoretical Foundation for Mechanistic Interpretability". In: *Journal of Machine Learning Research* 26.83 (2025), pp. 1–64. ISSN: 1533-7928. URL: http://jmlr.org/papers/v26/23-0058.html.

[Hub24] Martin Huber. "An introduction to causal discovery". In: *Swiss Journal of Economics and Statistics* 160.1 (Oct. 29, 2024), p. 14. ISSN: 2235-6282. DOI: 10.1186/s41937-024-00131-4. URL: https://sjes.springeropen.com/articles/10.1186/s41937-024-00131-4.

[JKZ19] Bart Jacobs, Aleks Kissinger, and Fabio Zanasi. "Causal Inference by String Diagram Surgery". In: *Foundations of Software Science and Computation Structures*. Ed. by Mikołaj Bojańczyk and Alex Simpson. Vol. 11425. Cham: Springer International Publishing, 2019, pp. 313–329. ISBN: 9783030171261 9783030171278. DOI: 10.1007/978-3-030-17127-8_18. URL: https://link.springer.com/10.1007/978-3-030-17127-8_18 (visited on 08/01/2025).

[Law62] William Lawvere. "The Category of Probabilistic Maps". In: (1962). URL: https://ncatlab.org/nlab/files/lawvereprobability1962.pdf.

[LT23] Robin Lorenz and Sean Tull. *Causal models in string diagrams.* arXiv:2304.07638. Apr. 2023. DOI: 10.48550/arXiv.2304.07638. URL: http://arxiv.org/abs/2304.07638.

[ML78] Saunders Mac Lane. *Categories for the Working Mathematician.* Vol. 5. Graduate Texts in Mathematics. New York, NY: Springer New York, 1978. ISBN: 9781441931238 9781475747218. DOI: 10.1007/978-1-4757-4721-8. URL: http://link.springer.com/10.1007/978-1-4757-4721-8.

[Par+24] Junhyung Park et al. *A Measure-Theoretic Axiomatisation of Causality.* arXiv:2305.17139. June 2024. DOI: 10.48550/arXiv.2305.17139. URL: http://arxiv.org/abs/2305.17139.

[Pet17] Jonas Peters. *Lectures on Causality: Jonas Peters, Part 1.* Mar. 2017. URL: https://www.youtube.com/watch?v=zvrcyqcN9Wo&list=PLW01hpWnEtbTcuY0a0jhZyanHX3GPImAy&index=1.

[PJS17] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms.* eng. Adaptive computation and machine learning. Cambridge, Mass: The MIT press, 2017. ISBN: 9780262037310.

[Tem+24] Adly Templeton et al. “Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet”. In: *Transformer Circuits Thread* (2024). URL: https://transformer-circuits.pub/2024/scaling-monosemanticity/index.html.

[Tri+17] Sofia Triantafillou et al. “Predicting Causal Relationships from Biological Data: Applying Automated Causal Discovery on Mass Cytometry Data of Human Immune Cells”. In: *Scientific Reports* 7.1 (Oct. 5, 2017), p. 12724. ISSN: 2045-2322. DOI: 10.1038/s41598-017-08582-x. URL: https://www.nature.com/articles/s41598-017-08582-x (visited on 08/01/2025).

[Tul+24] Sean Tull et al. *Towards Compositional Interpretability for XAI.* arXiv:2406.17583. June 2024. DOI: 10.48550/arXiv.2406.17583. URL: http://arxiv.org/abs/2406.17583.

[Çi11] Erhan Çinlar. *Probability and Stochastics.* en. Vol. 261. Graduate Texts in Mathematics. New York, NY: Springer New York, 2011. ISBN: 9780387878584 9780387878591. DOI: 10.1007/978-0-387-87859-1. URL: https://link.springer.com/10.1007/978-0-387-87859-1.