TECHNICAL REPORT
R-73
Apr, 2022

Causal Inference and Data Fusion: Towards an Accelerated Process of Scientific Discovery

Adèle H. Ribeiro and Elias Bareinboim

Causal Artificial Intelligence Laboratory
Department of Computer Science
Columbia University
{adele,eb}@cs.columbia.edu

Artificial intelligence (AI) often connotes futuristic speculations about how smart machines could become and whether they would eventually take over the planet. But far from the limelight of such extrapolations, a quiet AI revolution has already taken place, one that has profoundly transformed the way scientists look at the world, the language they use to interpret data, and the methods they use to assess cause and effect relationships in many social, environmental, and medical domains.

With the unprecedented accumulation of data and several breakthroughs in machine learning and deep neural networks over the past two decades, AI has proven capable of mastering and even outperforming humans at a variety of predictive tasks, including medical imaging diagnosis, strategy games (e.g. chess and Go), speech recognition, and driving cars. Despite numerous remarkable successes, most of these methods are still unable to provide causal explanations for their own decisions and behaviors [31, 32]. This contrast can be easily understood using the logic imposed by Pearl's Causal Hierarchy (PCH, for short) [34]. Prediction is a capability placed on the first layer of the PCH (known as associational), as predictive models can be learned from solely observational data. However, in general, there exist multiple models that induce the same observational distribution (and, therefore, have the same predictive power), but have different behaviors in terms of their causal and counterfactual explanations (capabilities placed on the second and third layers of the PCH, respectively). Therefore, there is no guarantee that a model with a high-predictive power is the true underlying causal model, also known as *structural causal model* (SCM, for short). In fact, one can prove that for *almost any* SCM, it is impossible to draw higher-layer inferences using only lower-layer information. This result has been formalized under the rubric of the *Causal Hierarchy Theorem* (CHT) [2, Thm. 1].

The CHT brings about an unfortunate yet enlightening lesson: the mere accumulation of data does not immediately translate into new insights about the underlying data-generating mechanisms, better predictions about the effects of new interventions, or estimates of retrospective counterfactuals. In fact, most observational datasets are collected under heterogeneous conditions - i.e., different populations, regimes, and sampling methods - which means they are plagued with various types of biases, including confounding, sampling selection, and cross-population (transportability) biases. Thus, all that can be claimed from these "messy" data collections are mostly statistical associations, not causation [6].

Researchers in the empirical fields are becoming increasingly aware that to take full advantage of this explosion of data and overcome some of the most pressing challenges in AI, such as explainability, generalizability, and fairness, current techniques must be enriched with two additional ingredients: the ability to distinguish causal relations from mere statistical correlations, and the ability to integrate knowledge and data from multiple, heterogeneous sources. Remarkably, the advent of graphical methods of causal and counterfactual inference [31, 6], and their recent connection with neural networks [36], have made it possible to tackle many of these challenging problems and have reignited hopes of constructing systems (software, machines, robots) capable of acting like human scientists and ultimately accelerating the process of scientific discovery.

It is well-understood in the literature that causal knowledge about the underlying system is necessary to perform causal inferences accurately, an idea popularized through Cartwright's motto [7]: "no causes-in, no causes-out." One major advantage of the modern theory of causation is its ability to infer causal effects in a non-parametric way, i.e., without requiring any knowledge about the functional form of the causal relationships or the distribution of the variables. It only requires knowledge about the underlying causal structure, which is transparently encoded in the form of a causal diagram. Notably, the recent introduction of cluster causal diagrams (or C-DAGs, for short) has allowed the partial specification of structural constraints, facilitating causal inferences in partially understood domains. Specifically, the language of C-DAGs provides a simple yet effective way to partially abstract a grouping of variables among which causal relationships are not fully understood while preserving consistency with the underlying causal system and the validity of identification of causal effects [1]. When knowledge is scarce, causal discovery algorithms can learn the equivalence class of the underlying causal diagram from data [35, 37, 16], which represents multiple models that explain equally well the evidence, and are therefore statistically indistinguishable.

Recent efforts based on the modern language of causation have culminated in a general framework for performing causal inference and data fusion [6], which brings substantial contributions to all steps of the process of scientific discovery. Based on the model assumptions encoded in the causal diagram (or its equivalence class) and an arbitrary collection of observational and experimental data, this framework establishes the conceptual basis for assessing identifiability and generalizability of specific causal queries about the reality being modeled. For instance, one can evaluate effects of previously unseen interventions [31, 3, 17, 18, 19, 29], reason about the effects of stochastic (and possibly imperfect) policies [8, 9], generalize causal knowledge to a target population [5, 33, 30, 10], recover from sample selection bias [4, 12, 13, 14], and derive new counterfactual explanations [11, 42]. Whenever an informative answer for the research question is computable, the corresponding effects can be estimated under various assumptions and with nice computational and statistical properties (e.g., double robustness, debiasedness) [20, 21, 23, 22]. On the other hand, if an answer *cannot* be achieve based on the available data and knowledge, one can yet efficiently design new strategies (e.g. measuring or experimenting over certain variables) to help get a positive answer in a next iteration of the process [25, 15, 24]. Taking these developments to the context of decision-making, one can now design more precise and surgical interventions [26, 27, 28], and further leverage observational data to accelerate the convergence of the exploratory process whenever decisions are not point identifiable [38, 39, 40, 41]. The tools that emerge from this framework have solved several long-standing challenges, including external validity, selection bias, meta-analysis, and transportability of experimental findings, which are pervasive in essentially every nontrivial instance of data analysis.

This mathematical framework, in practice, allows scientists to solve various challenges from first principles, which include reducing the cost of data collection and optimizing the design of experiments, predicting in domains with limited data and knowledge, and understanding the mechanisms underlying the phenomena under investigation. These issues are common challenges in a wide array of fields, including AI, machine learning, statistics, and the health and social sciences. In the area of robotics, for example, the results of this work can be used to endow intelligent systems with causal-generalization capabilities akin to the work that a human scientist conducts in a laboratory or field study. This means that a robot would be able to probe an environment more effectively, and then utilize the knowledge acquired to generalize to a new unexplored setting. Given the ubiquity of the data fusion problem across empirical disciplines, along with the generality and completeness of the current results, we believe that this new framework will be an essential tool for tackling the challenges presented to the next generation of data science research.

Acknowledgments

Adèle H. Ribeiro and Elias Bareinboim were supported in part by the Data Science Institute (Columbia University), the NSF, ONR, Amazon, JP Morgan, and The Alfred P. Sloan Foundation.

References

[1] Anand, T. V., Ribeiro, A. H., Tian, J., and Bareinboim, E. (2022). Effect identification in cluster causal diagrams. *arXiv preprint arXiv:2202.12263*.

[2] Bareinboim, E., Correa, J. D., Ibeling, D., and Icard, T. (2022). *On Pearl's Hierarchy and the Foundations of Causal Inference*, page 507–556. Association for Computing Machinery, NY, USA, 1st edition. Also, Columbia University, Causal AI Lab, Technical Report, R-60. https://causalai.net/r60.pdf.

[3] Bareinboim, E. and Pearl, J. (2012a). Causal inference by surrogate experiments: z-identifiability. In Freitas, N. and Murphy, K., editors, *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, pages 113–120, Catalina Island, CA. AUAI Press.

[4] Bareinboim, E. and Pearl, J. (2012b). Controlling selection bias in causal inference. In Lawrence, N. D. and Girolami, M., editors, *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pages 100–108, La Palma, Canary Islands. PMLR.

[5] Bareinboim, E. and Pearl, J. (2013). A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference*, 1(1):107–134.

[6] Bareinboim, E. and Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113:7345–7352.

[7] Cartwright, N. (1989). *Nature's Capacities and Their Measurement*. Clarendon Press, Oxford.

[8] Correa, J. and Bareinboim, E. (2019). From statistical transportability to estimating the effect of stochastic interventions. In Kraus, S., editor, *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, volume 28, pages 1661–1667, Macao, China. International Joint Conferences on Artificial Intelligence Organization.

[9] Correa, J. and Bareinboim, E. (2020a). A calculus for stochastic interventions: Causal effect identification and surrogate experiments. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY. AAAI Press.

[10] Correa, J. and Bareinboim, E. (2020b). General transportability of soft interventions: Completeness results. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 10902–10912, Vancouver, Canada. Curran Associates, Inc.

[11] Correa, J., Lee, S., and Bareinboim, E. (2021). Nested counterfactual identification from arbitrary surrogate experiments. In *Advances in Neural Information Processing Systems 34*. Curran Associates, Inc.

[12] Correa, J., Tian, J., and Bareinboim, E. (2018). Generalized adjustment under confounding and selection biases. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pages 6335–6342, New Orleans, LA. AAAI Press.

[13] Correa, J., Tian, J., and Bareinboim, E. (2019a). Adjustment criteria for generalizing experimental findings. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 1361–1369, Long Beach, CA. PMLR.

[14] Correa, J., Tian, J., and Bareinboim, E. (2019b). Identification of causal effects in the presence of selection bias. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, volume 33, pages 2744–2751, Honolulu, Hawaii. AAAI Press.

[15] Ghassami, A., Salehkaleybar, S., Kiyavash, N., and Bareinboim, E. (2018). Budgeted experiment design for causal structure learning. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 1724–1733, Stockholm, Sweden. PMLR.

[16] Jaber, A., Kocaoglu, M., Shanmugam, K., and Bareinboim, E. (2020). Causal discovery from soft interventions with unknown targets: Characterization and learning. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9551–9561, Vancouver, Canada. Curran Associates, Inc.

[17] Jaber, A., Zhang, J., and Bareinboim, E. (2018). Causal identification under markov equivalence. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence*, pages 978–987, Monterey, CA. AUAI Press.

[18] Jaber, A., Zhang, J., and Bareinboim, E. (2019a). Causal identification under markov equivalence: Completeness results. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 2981–2989, Long Beach, CA. PMLR.

[19] Jaber, A., Zhang, J., and Bareinboim, E. (2019b). Identification of conditional causal effects under markov equivalence. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E., and Garnett, R., editors, *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems*, volume 33, pages 11512–11520, Vancouver, Canada. Curran Associates, Inc.

[20] Jung, Y., Tian, J., and Bareinboim, E. (2020a). Estimating causal effects using weighting-based estimators. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY. AAAI Press.

[21] Jung, Y., Tian, J., and Bareinboim, E. (2020b). Learning causal effects via weighted empirical risk minimization. In *Advances in Neural Information Processing Systems*.

[22] Jung, Y., Tian, J., and Bareinboim, E. (2021a). Estimating identifiable causal effects on markov equivalence class through double machine learning. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5168–5179. PMLR.

[23] Jung, Y., Tian, J., and Bareinboim, E. (2021b). Estimating identifiable causal effects through double machine learning. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, Vancouver, Canada. AAAI Press.

[24] Kocaoglu, M., Jaber, A., Shanmugam, K., and Bareinboim, E. (2019). Characterization and learning of causal graphs with latent variables from soft interventions. In *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems*, volume 33, pages 14346–14356, Vancouver, Canada. Curran Associates, Inc.

[25] Kocaoglu, M., Shanmugam, K., and Bareinboim, E. (2017). Experimental design for learning causal graphs with latent variables. In *Advances in Neural Information Processing Systems*, pages 7018–7028. Curran Associates, Inc.

[26] Lee, S. and Bareinboim, E. (2018). Structural causal bandits: Where to intervene? In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 2568–2578, Montreal, Canada. Curran Associates, Inc.

[27] Lee, S. and Bareinboim, E. (2019). Structural causal bandits with non-manipulable variables. In

Proceedings of the 33rd AAAI Conference on Artificial Intelligence, volume 33, pages 4164–4172, Honolulu, Hawaii. AAAI Press.

[28] Lee, S. and Bareinboim, E. (2020). Characterizing Optimal Mixed Policies: Where to Intervene and What to Observe. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 8565–8576. Curran Associates, Inc.

[29] Lee, S., Correa, J., and Bareinboim, E. (2019). General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, volume 35, Tel Aviv, Israel. AUAI Press.

[30] Lee, S., Correa, J., and Bareinboim, E. (2020). Generalized transportability: Synthesis of experiments from heterogeneous domains. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY. AAAI Press.

[31] Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York.

[32] Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Commun. ACM*, 62(3):54–60.

[33] Pearl, J. and Bareinboim, E. (2014). External validity: From do-calculus to transportability across populations. *Statistical Science*, 29(4):579–595.

[34] Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books, Inc., New York, NY, USA, 1st edition.

[35] Spirtes, P., Glymour, C., and Scheines, R. (2001). *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition.

[36] Xia, K., Lee, K.-Z., Bengio, Y., and Bareinboim, E. (2021). The causal-neural connection: Expressiveness, learnability, and inference. *Advances in Neural Information Processing Systems*, 34.

[37] Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896.

[38] Zhang, J. and Bareinboim, E. (2017). Transfer learning in multi-armed bandits: A causal approach. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 1340–1346, Melbourne, Australia. International Joint Conferences on Artificial Intelligence Organization.

[39] Zhang, J. and Bareinboim, E. (2019). Near-optimal reinforcement learning in dynamic treatment regimes. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E., and Garnett, R., editors, *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems*, volume 33, pages 13401–13411, Vancouver, Canada. Curran Associates, Inc.

[40] Zhang, J. and Bareinboim, E. (2020). Designing optimal dynamic treatment regimes: A causal reinforcement learning approach. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119. PMLR.

[41] Zhang, J. and Bareinboim, E. (2021). Bounding causal effects on continuous outcome. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, Vancouver, Canada. AAAI Press.

[42] Zhang, J., Tian, J., and Bareinboim, E. (2021). Partial counterfactual identification from observational and experimental data. *arXiv preprint arXiv:2110.05690*.