

MIT Press • Open Encyclopedia of Cognitive Science

Causal Learning

Alison Gopnik¹

¹University of California at Berkeley

MIT Press

Published on: Jul 24, 2024

DOI: <https://doi.org/10.21428/e2759450.02bf2682>

License: [Creative Commons Attribution-NonCommercial 4.0 International License \(CC-BY-NC 4.0\)](#).

Learning about the causal structure of the world is a fundamental problem for human cognition. Over the past 20 years, cognitive scientists have applied advances in our understanding of causation in philosophy and computer science, particularly within the Causal Bayes Net formalism, to understand human causal learning. The formalism specifies a probabilistic generative model that describes the causal relations between variables. The central idea is that if X causes Y, acting to change the value of X will change the value of Y. Empirical research has shown that human children are remarkably adept at causal learning and that they learn in a way that mirrors the formalism. Preschool children can use information about the conditional probability of variables and the outcomes of actions on those variables to infer the causal structure of the physical, biological, and psychological world; discover unobserved variables; and learn high-level abstract features of causal systems. Children do this by observing the outcomes of other people's actions, listening to causal language, and exploring and experimenting themselves in their everyday play.

History

Going back to [Hume \(1739\)](#), philosophers have worried about how we can learn that X causes Y, when all we directly observe is whether X is associated with Y. Some suggested that causation just *is* association; others argued that causal relations require a particular spatio-temporal path from cause to effect. Around 2000, philosophers of science James Woodward and Clark Glymour and computer scientist Judea Pearl articulated a formal account of causal knowledge and learning ([Glymour, 2001](#); [Spirtes et al., 2001](#); [Woodward, 2005, 2021](#); [Pearl, 2000](#); [Pearl & Mackenzie, 2018](#)).

In parallel, developmental psychologists articulated the “theory theory”: the idea that children's cognitive development and learning is like theory formation in science [see [Cognitive Development](#)] ([Gopnik, 2012](#); [Carey, 2009](#); [Wellman & Gelman, 1998](#)). Causal learning plays a central role in theory formation, so a research program emerged to test whether children would infer the right causal structures from data about statistics and interventions in the way that scientists do and that the formalism prescribed. This turned out to be correct to a remarkable degree. By age four, children can infer the causal structure of the physical, biological, and psychological world; discover unobserved variables; and learn high-level causal *overhypotheses* in this way ([Gopnik & Wellman, 2012](#); [Gopnik & Gopnik, 2024](#)). Also, in parallel, developmentalists found that even infants could detect some particular causal relationships, such as relations of movement and collision among objects and relations between agents' actions and their goals ([Carey, 2009](#); [Spelke, 2022](#)).

Core concepts

Intervention

The *interventionist* definition of causation is that a variable X is causally related to a variable Y if and only if an intervention to change the value of X will change the value of Y. This is why a controlled experiment

is the gold standard for causal inference in science. It contrasts with association or prediction. Yellow nicotine-stained fingers and smoking might both be associated with lung cancer, and you might predict that someone with yellow fingers would be more likely to get lung cancer. But intervening to wash your hands would not affect the probability of the disease, although stopping smoking would.

Bayesian networks

Causal relations are often described in terms of directed acyclic graphs that systematically connect causal variables, called *Bayes nets*. The graph can be used to systematically generate the conditional probabilities of variables and the outcomes of interventions on those variables. In principle, it is then possible to use Bayesian methods to learn the graph from the data. The Bayes net formalism was part of a broader movement toward Bayesian probabilistic models of cognition [see [Bayesianism](#); [Bayesian Models of Cognition](#)], involving many types of probabilistic generative models and including hierarchical relations between such models ([Tenenbaum et al., 2011](#); [Ullman & Tenenbaum, 2020](#)).

Overhypotheses

A higher-order causal overhypothesis or *framework theory* could constrain lower-level models ([Griffiths & Tenenbaum, 2009](#)). For example, if you believe that human actions are caused by beliefs and desires [see [Theory of Mind](#)], you can restrict lower-level causal explanations for a particular action to Bayes nets that include these variables.

Search

The Bayesian hypothesis testing framework has a central dilemma: How does a learner search through all the possible causal hypotheses and decide which ones to test? There have been attempts to answer this question using “sampling” methods, and children may use these methods [see [Markov Chain Monte Carlo](#)] ([Bonawitz et al., 2014](#)). But the search problem is still intractable and hence an open area of research.

Active learning and exploration

Children actively and spontaneously intervene on the world to get data to solve causal problems ([Schulz, 2012](#)). This active learning is similar to experimentation in science and to exploration in reinforcement learning.

Social causal learning

Children learn causal relationships both through active learning and exploration but also through observations of others and through language [see [Social Learning](#)] ([Sobel & Kushnir, 2013](#); [Meltzoff et al., 2012](#)).

Questions, controversies, and new developments

How is general causal learning related to more specific types of causal knowledge in *intuitive physics* or *intuitive psychology*, sometimes called *core knowledge* ([Carey, 2009](#); [Spelke, 2022](#)), which appear to be in place very early in development? Are these separate systems or are they integrated? And if so, how?

Different caregivers and cultures may transmit different kinds of causal knowledge through language and various forms of teaching. How these linguistic and cultural mechanisms work is still an unsolved problem.

Broader connections

One important new direction concerns the relation between causal learning and the machine learning methods that have recently transformed computer science and AI. Even very large language models appear to have particular difficulty with causal inference and learning tasks [see [Large Language Models](#)]. Is there a way to combine modern machine learning and the causal Bayes net approach? Modifying reinforcement learning techniques, which, like causal learning, involve interventions on the world, might be especially promising.

A second set of broader connections concerns the biological and evolutionary origins of causal learning. On one hand, understanding the effects of your actions on the world, a primordial kind of causal learning, is foundational to even the first intelligences that evolved in the Cambrian, the first geological period of the Paleozoic Era. On the other hand, the comparative animal cognition literature suggests that even other primates' causal learning abilities are limited compared to human learning. Nonhuman animals appear to learn in a way that is less general—restricted to particular ecologically relevant contexts—and less impersonal—restricted to learning the effects of your own actions rather than the effects of others' actions or understanding causal relations that are not directly the outcome of actions ([Goddu & Gopnik, 2024](#)).

Acknowledgements

The author acknowledges funding and support from DARPA grant 047498-002, John Templeton Foundation grant 61475, Templeton World Charity Foundation TWCF0434, and the Department of Defense Office of Naval Research Multidisciplinary University Initiative.

Further reading

- Goddu, M.K., Gopnik, A. (2024). The development of human causal learning and reasoning. *Nature Reviews Psychology*, 3(5), 319–339. <https://doi.org/10.1038/s44159-024-00300-5>
- Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic books.
- Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *Annual Review of Developmental Psychology*, 2, 533–558. <https://doi.org/10.1146/annurev-devpsych-121318-084833>

- Woodward, J. (2021). *Causation with a human face: Normative theory and descriptive psychology*. Oxford University Press.

References

- Bonawitz, E., Denison, S., Griffiths, T. L., & Gopnik, A. (2014). Probabilistic models, learning algorithms, and response variability: Sampling in cognitive development. *Trends in Cognitive Sciences*, 18, 497–500. Advance online publication. <https://doi.org/10.1016/j.tics.2014.06.006>

↩

- Carey, S. (2009). *The origin of concepts*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195367638.001.0001>

↩

- Glymour, C. N. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. MIT press. <https://doi.org/10.7551/mitpress/4638.001.0001>

↩

- Goddu, M.K., Gopnik, A. (2024). The development of human causal learning and reasoning. *Nature Reviews Psychology*, 3(5), 319–339. <https://doi.org/10.1038/s44159-024-00300-5>

↩

- Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, 337(6102), 1623–1627. <https://doi.org/10.1126/science.1223416>

↩

- Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, 138(6), 1085–1108. <https://doi.org/10.1037/a0028044>

↩

- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116(4), 661–716. <https://doi.org/10.1037/a0017201>

↩

- Hume, D. (1739). *A treatise of human nature*. Clarendon Press.

↩

- Meltzoff, A. N., Waismeyer, A., & Gopnik, A. (2012). Learning about causes from people: Observational causal learning in 24-month-old infants. *Developmental Psychology*, 48(5), 1215.

↩

- Pearl, J. (2000). *Causality*. Cambridge University Press.

↩

- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

↩

- Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, 16, 382–389. <https://doi.org/10.1016/j.tics.2012.06.004>

↩

- Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, 120(4), 779–797. <https://doi.org/10.1037/a0034191>

↩

- Spelke, E. S. (2022). *What babies know: Core knowledge and composition* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/oso/9780190618247.001.0001>

↩

- Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search*. MIT press. <https://doi.org/10.7551/mitpress/1754.001.0001>

↩

- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022), 1279–1285. <https://doi.org/10.1126/science.1192788>

↩

- Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *Annual Review of Developmental Psychology*, 2, 533–558. <https://doi.org/10.1146/annurev-devpsych-121318-084833>

↩

- Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In W. Damon (Ed.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 523–573). John Wiley & Sons, Inc.

↩

- Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford University Press.

↩

- Woodward, J. (2021). *Causation with a human face: Normative theory and descriptive psychology*. Oxford University Press.

↩