David Danks

Curriculum Vitae

Contact Information

Halicioğlu Data Science Institute (HDSI) (858) 246-4548
9500 Gilman Dr MC 0555 ddanks@ucsd.edu
University of California, San Diego http://www.daviddanks.org
La Jolla, CA 92093 *Country of citizenship*: United States

Academic Career

University of California, San Diego
- Professor of Data Science, Philosophy, & Policy (joint appointment) 2021 -
- Faculty Council Chair, HDSI 2023 -
- *Affiliate & adjunct appointments:*
  - Department of Computer Science & Engineering, UCSD 2021 -
  - Center for Philosophy of Science, University of Pittsburgh 2005 -

Past appointments:

Carnegie Mellon University
- Louis Leon (L. L.) Thurstone Professor of Philosophy & Psychology 2016 - 2021
- Head, Department of Philosophy 2014 - 2021
- Professor of Philosophy & Psychology 2014 - 2016
- Associate Professor of Philosophy & Psychology 2008 - 2014
- Assistant Professor of Philosophy 2003 - 2008
- *Affiliate status while at CMU*:
  - Carnegie Mellon Neuroscience Institute 2019 - 2021
  - H. John Heinz III College of Information Systems and Public Policy 2017 - 2021
  - Center for the Neural Basis of Cognition 2014 - 2021
  - Department of History & Philosophy of Science, Univ. of Pittsburgh 2014 - 2021

Center for Advanced Study of Language, University of Maryland
- Adjunct member 2008 - 2019

Florida Institute for Human & Machine Cognition (UARI with Univ. of Florida system)
- Research Scientist 2001 - 2012

Philosophy Dept., Colorado College
- Visiting Assistant Professor (2 courses) 2002 - 2003

Education

University of California, San Diego (La Jolla, Calif.)
- Philosophy Department (Ph.D., 6/2001; M.A., 12/1999) 1996 - 2001
- Dissertation: *The Epistemology of Causal Judgment*

Carnegie Mellon University (Pittsburgh, Pa.)
- Logic, Computation, and Methodology (visiting graduate student) 1998 - 1999

Princeton University (Princeton, N.J.)
- Major: Philosophy (A.B. *cum laude*, 5/1996) 1992 - 1996

Publications *[links to most papers available on webpage]*

Books:

Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. Cambridge, MA: The MIT Press.

Reviews: *Notre Dame Philosophical Reviews* by Steven Horst (2014); *Choice* by Steven Horst (March, 2015); *PsycCRITIQUES* by Donald MacGregor (March 23, 2015); *Zagadnienia Naukoznawstwa* by Pawel Kawalec (2015) [review in Polish]; *Philosophical Psychology* by Christopher Burr (2016)

Danks, D., & Ippoliti, E. (Eds.) (2018). *Building theories: Heuristics and hypotheses in science*. Berlin: Springer-Verlag.

Journal articles and book chapters:

[75] Amico-Korby, D., Harrell, M., & Danks, D. (in press). Building epistemically healthier platforms. *Episteme*.

[74] Danks, D., & Harrell, M. (in press). Chaos, causation, and describing dynamics. In C. K. Waters (Ed.), *Causal reasoning in biology*. Minneapolis: University of Minnesota Press.

[73] Amico-Korby, D., Harrell, M., & Danks, D. (2024). Environmental epistemology. *Synthese, 203*(81), 24 pages.

[72] Cusimano, C., Zorrilla, N., Danks, D., & Lombrozo, T. (2024). Psychological freedom, rationality, and the naïve theory of reasoning. *Journal of Experimental Psychology: General, 153*(3), 837-863.

[71] Sloane, M., Danks, D., & Moss, E. (2024). Tackling AI *hyping. AI & Ethics*.

[70] Swaminathan, N., & Danks, D. (2024). Governing ethical gaps in distributed AI development. *Digital Society*.

[69] Bakirtzis, G., Carr, S., Danks, D., & Topcu, U. (2023). Dynamic certification for autonomous systems. *Communications of the ACM, 66*(9): 64-72.

[68] Danks, D. (2023). Pragmatism and the challenge of scientific (dis)unification. In H. K. Andersen & S. D Mitchell (Eds.), *The pragmatist challenge: Pragmatist metaphysics for philosophy of science* (pp. 160-179) Oxford: Oxford University Press.

[67] Danks, D., & Davis, I. (2023). Causal inference in cognitive neuroscience. *WIREs Cognitive Science,* e1650. doi:10.1002/wcs.1650

[66] Hagendorff, T., & Danks, D. (2023). Ethical and methodological challenges in building morally informed AI systems. *AI & Ethics, 3*, 553-566. https://doi.org/10.1007/s43681-022-00188-y

[65] Schnetz, M., Danks, D., & Mahajan, A. (2023). Preoperative identification of patient-dependent blood pressure targets associated with low risk of intraoperative hypotension during noncardiac surgery. *Anesthesia & Analgesia, 136*(2): 194-203.

[64] Trusilo, D., & Danks, D. (2023). Artificial intelligence and humanitarian obligations. *Ethics and Information Technology, 25*: article 12 (5 pages).

[63] Danks, D. (2022). Governance via explainability. In J. Bullock (Ed.), *The Oxford handbook of AI governance*. Oxford: Oxford University Press.

[62] Danks, D., & Dinh, P. N. (2022). Causal perception and causal inference: An integrated account. In P. Willemsen & A. Wiegmann (Eds.), *Advances in Experimental Philosophy of Causation* (pp. 81-100). New York, NY: Bloomsbury.

[61] Danks, D., & Trusilo, D. (2022). The challenge of ethical interoperability. *Digital Society, 1*: article 11.

[60] Fazelpour, S., Lipton, Z. C., & Danks, D. (2022). Algorithmic fairness and the situated dynamics of justice. *Canadian Journal of Philosophy, 52*(1): 44-60.

[59] McCaffrey, J., & Danks, D. (2022). Mixtures and psychological inference with resting state fMRI. *British Journal for the Philosophy of Science, 73*(3): 583-611. doi:10.1093/bjps/axx053

[58] National Academies of Sciences, Engineering, and Medicine. (2022). *Fostering responsible computing research: Foundations and practices*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26507.

[57] National Academies of Sciences, Engineering, and Medicine. (2022). *Ontologies in the behavioral sciences: Accelerating research and the spread of knowledge*. Washington, DC: The National Academies Press.

[56] Danks, D. (2021). Digital ethics as translational ethics. In I. Vasiliu-Feltes & J. Thomason (Eds.), *Applied ethics in a digital world* (pp. 1-15). IGI Global.

[55] Dinh, P. N., & Danks, D. (2021). Causal pluralism in philosophy: Empirical challenges and alternative proposals. *Philosophy of Science, 88*(5): 761-772.

[54] Falco, G., Shneiderman, B., & 18 additional authors (alphabetical) including Danks, D. (2021). Governing AI safety through independent audits. *Nature Machine Intelligence, 3*, 566-571.

[53] Fazelpour, S., & Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophy Compass, 16*(8), e12760.

[52] Lütge, C., Poszler, F., Acosta, A. J., Danks, D., Gottehrer, G., Mihet-Popa, L., & Naseer, A. (2021). AI4People: Ethical guidelines for the automotive sector – fundamental requirements and practical recommendations. *International Journal of Technoethics, 12*(1), 101-125.

[51] Montague, E., Day, T. E., Barry, D., Brumm, M., McAdie, A., Cooper, A. B., Wignall, J., Erdman, S., Núñez, D., Diekema, D., & Danks, D. (2021). The case for information fiduciaries: The implementation of a data ethics checklist at Seattle Children's Hospital. *Journal of the American Medical Informatics Association, 28*, 650-652. doi:10.1093/jamia/ocaa307

[50] Danks, D. (2019). Probabilistic models. In M. Colombo & M. Sprevak (Eds.), *Routledge handbook of the computational mind* (pp. 149-158). New York: Routledge.

[49] Danks, D. (2019). Safe-and-substantive perspectivism. In M. Massimi & C. D. McCoy (Eds.), *Understanding perspectivism: Scientific challenges and methodological prospects* (pp. 127-140). New York: Routledge.

[48] Danks, D., & Plis, S. M. (2019). Amalgamating evidence of dynamics. *Synthese, 196*(8), 3213-3230.

[47] Schnetz, M. P., Hochheiser, H. S., Danks, D. J., Landsittel, D. P., Vogt, K. M., Ibinson, J. W., Whitehurst, S. L., McDermott, S. P., Duque, M. G., & Kaynar, A. M. (2019). The Triple Variable Index combines information generated over time from common monitoring variables to identify patients expressing distinct patterns of intraoperative physiology. *BMC Medical Research Methodology, 19*(1): 17 pages. doi:10.1186/s12874-019-0660-9

[46] Danks, D. (2018). LPCD framework: Analytical tool or psychological model? [Commentary] *Behavioral and Brain Sciences, 41*, E230. doi:10.1017/S0140525X18001383

[45] Danks, D. (2018). Privileged (default) causal cognition: A mathematical analysis. *Frontiers in Psychology, 9*: 498. doi:10.3389/fpsyg.2018.00498

[44] Danks, D. (2018). Richer than reduction. In D. Danks & E. Ippoliti (Eds.), *Building theories: Heuristics and hypotheses in science* (pp. 45-61). Berlin: Springer-Verlag.

[43] Malinsky, D., & Danks, D. (2018). Causal discovery algorithms: A practical guide. *Philosophy Compass, 13*, e12470. doi:10.1111/phc3.12470

[42] Roff, H. M., & Danks, D. (2018). "Trust but Verify": The difficulty of trusting autonomous weapons systems. *Journal of Military Ethics, 17*, 2-20.

[41] Broger, T., Roy, R. B., Filomena, A., Greef, C. H., Rimmele, S., Havumaki, J., Danks, D., & 11

additional authors. (2017). Diagnostic performance of tuberculosis-specific IgG antibody profiles in patients with presumptive TB from two continents. *Clinical Infectious Diseases, 64*, 947-955. doi:10.1093/cid/cix023

[40] Danks, D. (2017). Singular causation. In M. R. Waldmann (Ed.), *Oxford handbook of causal reasoning* (pp. 201-215). Oxford: Oxford University Press.

[39] Danks, D., & London, A. J. (2017). Regulating autonomous systems: Beyond standards. *Intelligent Systems, 32(1)*, 88-91.

[38] Hyttinen, A., Plis, S., Järvisalo, M., Eberhardt, F., & Danks, D. (2017). A constraint optimization approach to causal discovery from subsampled time series data. *International Journal of Approximate Reasoning, 90*, 208-225.

[37] Danks, D. (2016). Causal search, causal modeling, and the folk. In J. Sytsma & J. W. Buckwalter (Eds.), *Blackwell companion to experimental philosophy* (pp. 463-471). Oxford: Wiley Blackwell.

[36] Danks, D., & Danks, J. H. (2016). Beyond machines: Humans in cyber operations, espionage, and conflict. In F. Allhoff, A. Henschke, & B. J. Strawser (Eds.), *Binary bullets: The ethics of cyberwarfare* (pp. 177-197). Oxford: Oxford University Press.

[35] Wellen, S., & Danks, D. (2016). Adaptively rational learning. *Minds & Machines, 26*(1), 87-102. DOI: 10.1007/s11023-015-9370-1

[34] Danks, D. (2015). Goal-dependence in (scientific) ontology. *Synthese, 192*, 3601-23616. DOI: 10.1007/s11229-014-0649-1

[33] Danks, D. (2014). Learning. In K. Frankish & W. M. Ramsey (Eds.), *Cambridge handbook to artificial intelligence* (pp. 151-167). Cambridge: Cambridge University Press.

[32] Danks, D. (2014). A modern Pascal's wager for mass electronic surveillance. *Telos, 169*, 155-161.

[31] Danks, D., Rose, D., & Machery, E. (2014). Demoralizing causation. *Philosophical Studies, 171*(2), 251-277.

[30] Kummerfeld, E., & Danks, D. (2014). Model change and reliability in scientific inference. *Synthese, 191*(12), 2673-2693.

[29] Danks, D. (2013). Functions and cognitive bases for the concept of actual causation. *Erkenntnis, 78*(1), 111-128. DOI: 10.1007/s10670-013-9439-2

[28] Danks, D., & Danks, J. H. (2013). The moral permissibility of automated responses during cyberwarfare. *Journal of Military Ethics, 12(1)*, 18-33.

[27] Mayo-Wilson, C., Zollman, K. J. S., & Danks, D. (2013). Wisdom of crowds vs. groupthink: Learning in groups and in isolation. *International Journal of Game Theory, 42(3)*, 695-723.

[26] Rose, D., & Danks, D. (2013). In defense of a broad conception of experimental philosophy. *Metaphilosophy, 44(4)*, 512-532.

[25] Danks, D. (2012). Human causal learning. In N. Seel (Ed.), *Encyclopedia of the sciences of learning*. Springer.

[24] Rose, D., & Danks, D. (2012). Causation: Empirical trends and future directions. *Philosophy Compass, 7(9)*, 643-653.

[23] Danks, D., & Eberhardt, F. (2011). Integration in both directions: The need for an account of algorithmic rationality [Commentary]. *Brain & Behavioral Sciences, 34*, 197.

[22] Eberhardt, F., & Danks, D. (2011). Confirmation in the cognitive sciences: The problematic case of Bayesian models. *Minds and Machines, 21(3)*, 389-410.

[21] Mayo-Wilson, C., Zollman, K. J. S., & Danks, D. (2011). The independence thesis: When individual

and social epistemology diverge. *Philosophy of Science, 78(4)*, 653-677.

[20] Danks, D. (2010). Not different kinds, just special cases [Commentary]. *Behavioral and Brain Sciences, 33(2/3),* 208-209.

[19] Danks, D., Fancsali, S., Glymour, C., & Scheines, R. (2010). Comorbid science? [Commentary]. *Behavioral and Brain Sciences, 33(2/3),* 153-155.

[18] Danks, D., & Rose, D. (2010). Diversity in representations, uniformity in learning [Commentary]. *Behavioral and Brain Sciences, 33*, 90-91.

[17] Glymour, C., Danks, D., Glymour, B., Eberhardt, F., Ramsey, J., Scheines, R., Spirtes, P., Teng, C. M., & Zhang, J. (2010). Actual causation: A stone soup essay. *Synthese, 175(2)*, 169-192.

[16] Ramapriyan, H., Isaac, D., Yang, W., Bonnlander, B., & Danks, D. (2010). An intelligent archive testbed incorporating data mining. In L. Di & H. K. Ramapriyan (Eds.), *Standard-based data and information systems for earth observations* (pp. 165-188). Berlin: Springer-Verlag.

[15] Danks, D. (2009). The psychology of causal perception and reasoning. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation* (pp. 447-470). Oxford: Oxford University Press.

[14] Danks, D., & Eberhardt, F. (2009). Conceptual problems in statistics, testing and experimentation. In J. Symons & F. Calvo (Eds.), *Routledge companion to the philosophy of psychology* (pp. 214-230). New York: Routledge. 2nd edition (in press). [includes new section on the "replication crisis"]

[13] Danks, D., & Eberhardt, F. (2009). Explaining norms and norms explained [Commentary]. *Behavioral and Brain Sciences, 32* (1), 86-87.

[12] Wimberly, F., Danks, D., Glymour, C., & Chu, T. (2009). Problems for structure learning: Aggregation and computational complexity. In S. Das, S. M. Welch, D. Caragea, & W. H. Hsu (Eds.), *Computational methodologies in gene regulatory networks* (pp. 310-332). Hershey, PA: IGI Global Publishing.

[11] Danks, D. (2008). Rational analyses, instrumentalism, and implementations. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 59-75). Oxford: Oxford University Press.

[10] Jantzen, B., & Danks, D. (2008). Biological codes and topological causation. *Philosophy of Science, 75*, 259-277.

[9] Townsend, K. A., Wollstein, G., Danks, D., Sung, K. R., Ishikawa, H., Kagemann, L., Gabriele, M. L., & Schuman, J. S. (2008). Heidelberg Retina Tomography III machine learning classifiers for glaucoma detection. *British Journal of Ophthalmology, 92*, 814-818.

[8] Danks, D. (2007). Causal learning from observations and manipulations. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 359-388). New York: Lawrence Erlbaum Associates.

[7] Danks, D. (2007). Theory unification and graphical models in human categorization. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 173-189). Oxford: Oxford University Press.

[6] Glymour, C., & Danks, D. (2007). Reasons as causes in Bayesian epistemology. *Journal of Philosophy, 104(9)*, 464-474.

[5] Scheines, R., Easterday, M., & Danks, D. (2007). Teaching the normative theory of causal reasoning. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 119-138). Oxford: Oxford University Press.

[4] Danks, D. (2005). Scientific coherence and the fusion of experimental results. *The British Journal for the Philosophy of Science*, *56*, 791-807.

[3] Danks, D. (2005). The supposed competition between theories of human causal inference.

Philosophical Psychology, *18* (2), 259-272.

[2] Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111* (1), 3-32.

[1] Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, *47*, 109-121.

Peer-reviewed conference proceedings:

[48] Chien, J., & Danks, D. (2024). Beyond behaviorist representational harms: A plan for measurement and mitigation. *FAccT 2024*.

[47] Kulzhabayeva, D., Danks, D., & Williams, J. (2024). Dynamics of causal attribution. *Proceedings of the 46th annual meeting of the Cognitive Science Society*.

[46] Bontula, A., Danks, D., & Fitter, N. T. (2023). The ambiguity of robot rights. In A. A. Ali, J.-J. Cabibihan, N. Meskin, S. Rossi, W. Jiang, H. He, & S. S. Ge (Eds.), *Proceedings of ICSR 2023* (pp. 204-215). Springer.

[45] Abavisani, M., Danks, D., & Plis, S. (2023). GRACE-C: Generalized rate agnostic causal estimation via constraints. *Proceedings of ICLR*. **notable top 25% awardee**

[44] Dinh, P., & Danks, D. (2023). Expectations of determinism underlie domain effects on adult causal learning. *Proceedings of the 45th annual meeting of the cognitive science society*.

[43] Solovyeva, K., Danks, D., Abavisani, M., & Plis, S. (2023). Causal learning through deliberate undersampling. *Proceedings of CLeaR conference*.

[42] Dinh, P. D., & Danks, D. (2022). Expectations of causal determinism in causal learning. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual meeting of the cognitive science society*. Austin, TX: Cognitive Science Society.

[41] Fogliato, R., Fazelpour, S., Gupta, S., Lipton, Z., & Danks, D. (2022). Homophily and incentive effects in use of algorithms. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual meeting of the cognitive science society*. Austin, TX: Cognitive Science Society.

[40] Cusimano, C., Zorrilla, N., Danks, D., & Lombrozo, T. (2021). Reason-based constraint in theory of mind. In *Proceedings of the 43rd annual conference of the cognitive science society*.

[39] Golden, P., & Danks, D. (2021). Ethical obligations to provide novelty. In *Proceedings of the 2021 AAAI/ACM Conference on Artificial Intelligence, Ethics, & Society*.

[38] Johnston, L., Hillman, N., & Danks, D. (2021). Individual differences in causal learning. In *Proceedings of the 43rd annual conference of the cognitive science society*.

[37] Lu, J., Lee, D., Kim, T. W., & Danks, D. (2020). Good explanation for algorithmic transparency. In A. Markham, J. Powles, T. Walsh, & A. L. Washington (Eds.), *Proceedings of the 2020 AAAI/ACM Conference on Artificial Intelligence, Ethics, & Society*. New York: ACM.

[36] Zhou, Y., & Danks, D. (2020). Different "intelligibility" for different folks. In A. Markham, J. Powles, T. Walsh, & A. L. Washington (Eds.), *Proceedings of the 2020 AAAI/ACM Conference on Artificial Intelligence, Ethics, & Society* (pp. 194-200). New York: ACM.

[35] Danks, D. (2019). The value of trustworthy AI. In *Proceedings of the 2019 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*.

[34] Geary, T., & Danks, D. (2019). Balancing the benefits of autonomous vehicles. In *Proceedings of the 2019 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*.

[33] Parker, J., & Danks, D. (2019). How technological advances can reveal rights. In *Proceedings of the 2019 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*.

[32] LaRosa, E., & Danks, D. (2018). Impacts on trust of healthcare AI. In *Proceedings of the 2018 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*. doi:10.1145/3278721.3278771

[31] London, A. J., & Danks, D. (2018). Regulating autonomous vehicles: A policy proposal. In *Proceedings of the 2018 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*. doi:10.1145/3278721.3278763

[30] Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. In C. Sierra (Ed.), *Proceedings of the 26th International Joint Conference on Artificial Intelligence* (pp. 4691-4697).

[29] Kazman, R., Stoddard, R., Danks, D., & Cai, Y. (2017). Causal modeling, discovery, & inference for software engineering. In *Proceedings of 39th International Conference on Software Engineering (ICSE 2017)* (pp. 172-174). Piscataway, NJ: IEEE Press.

[28] Hyttinen, A., Plis, S., Järvisalo, M., Eberhardt, F., & Danks, D. (2016). Causal discovery from subsampled time series data by constraint optimization. In A. Antonucci, G. Corani, & C. P. de Campos (Eds.), *JMLR Workshop & Conference Proceedings* (vol. 52): *Proceedings of the 8th International Conference on Probabilistic Graphical Models* (pp. 216-227).

[27] Plis*, S., Danks*, D., Freeman, C., & Calhoun, V. (2015). Rate-agnostic (causal) structure learning. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems 28* (pp. 3303-3311). La Jolla, CA: The NIPS Foundation. [*first two authors contributed equally]

[26] Plis*, S., Danks*, D., & Yang, J. (2015). Mesochronal structure learning. In M. Meila & T. Heskes (Eds.), *Uncertainty in artificial intelligence 31 (UAI-2015)* (pp. 702-711). Corvallis, OR: AUAI Press. [*first two authors contributed equally]

[25] Danks*, D., & Plis*, S. (2014). Learning causal structure from undersampled time series. In *JMLR: Workshop and Conference Proceedings*. [*authors contributed equally]

[24] Wellen, S., & Danks, D. (2014). Learning with a purpose: The influence of goals. In P. Bello, M Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (pp. 1766-1771). Austin, TX: Cognitive Science Society.

[23] Kummerfeld, E., & Danks, D. (2013). Tracking time-varying graphical structure. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K.Q. Weinberger (Eds.), *Advances in neural information processing systems 26* (pp. 1205-1213). La Jolla, CA: The NIPS Foundation.

[22] Danks, D. (2013). Moving from levels & reduction to dimensions & constraints. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 2124-2129). Austin, TX: Cognitive Science Society.

[21] Nevins, J. E., Danks, D., Wollstein, G., Ishikawa, H., Kagemann, L., Sigal, I. A., & Schuman, J. S. (2013). Machine classifier clustering of ocular structure measurements poorly corresponds with longitudinal functional performance in glaucoma. *Association for Research in Vision and Ophthalmology (ARVO) 2013*.

[20] Wellen, S., & Danks, D. (2012). Actor-observer asymmetries in judgments of intentional actions. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 2523-2528). Austin, TX: Cognitive Science Society.

[19] Wellen, S., & Danks, D. (2012). Learning causal structure through local prediction-error learning. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 2529-2534). Austin, TX: Cognitive Science Society.

[18] Lally, D. R., Wollstein, G., Danks, D., Ishikawa, H., Kagemann, L., & Schuman, J.S. (2009). Combining OCT, HRT and GDx through machine learning classifiers for glaucoma detection. *Association for Research in Vision and Ophthalmology (ARVO) 2009*.

[17] Tillman, R. E., Danks, D., & Glymour, C. (2008). Integrating locally learned causal structures with overlapping variables. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems 21* (pp. 1665-1672). La Jolla, CA: The NIPS Foundation.

[16] Nichols, W., & Danks, D. (2007). Decision making using learned causal structures. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the cognitive science society* (pp. 1343-1348). Austin, TX: Cognitive Science Society.

[15] Townsend, K. A., Wollstein, G., Danks, D., Sung, K., Ishikawa H., Kagemann, L., Gabriele, M. L., & Schuman, J. S. (2007). Heidelberg Retina Tomography 3 machine learning classifiers for glaucoma detection. *Association for Research in Vision and Ophthalmology (ARVO) 2007*.

[14] Zhu, H., & Danks, D. (2007). Task influences on category learning. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the cognitive science society* (pp. 1677-1682). Austin, TX: Cognitive Science Society.

[13] Danks, D. (2006). (Not) learning a complex (but learnable) category. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual meeting of the cognitive science society* (pp. 1186-1191). Mahwah, NJ: Lawrence Erlbaum Associates.

[12] Danks, D., & Schwartz, S. (2006). Effects of causal strength on learning from biased sequences. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual meeting of the cognitive science society* (pp. 1180-1185). Mahwah, NJ: Lawrence Erlbaum Associates.

[11] Ramapriyan, H. K., Isaac, D., Yang, W., Bonnlander, B., & Danks, D. (2006). An intelligent archive testbed incorporating data mining lessons and observations. In *Proceedings of the IEEE geoscience and remote sensing symposium* (pp. 3482-3485).

[10] Danks, D., & Schwartz, S. (2005). Causal learning from biased sequences. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual meeting of the cognitive science society* (pp. 542-547). Mahwah, NJ: Lawrence Erlbaum Associates.

[9] Bunch, L., Breedy, M., Bradshaw, J. M., Carvalho, M., Danks, D., & Suri, N. (2004). Flexible automated monitoring and notification for complex processes. In F.-Y. Wang (Ed.), *Proceedings of the IEEE international conference on networking, sensing, and control* (pp. 443-448). Tucson, AZ.

[8] Danks, D. (2004). Constraint-based human causal learning. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.), *Proceedings of the 6th international conference on cognitive modeling (ICCM-2004)* (pp. 342-343). Mahwah, NJ: Lawrence Erlbaum Associates.

[7] Danks, D., Glymour, C., & Spirtes, P. (2003). The computational and experimental complexity of gene perturbations for regulatory network search. In W. H. Hsu, R. Joehanes, and C. D. Page (Eds.), *Proceedings of IJCAI workshop on learning graphical models for computational genomics* (pp. 22-31).

[6] Danks, D., Griffiths, T. L., & Tenenbaum, J. B. (2003). Dynamical causal learning. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 67-74). Cambridge, MA: MIT Press.

[5] Hewett, R., & Danks, D. (2003). Integration of learning with probabilistic and compact relational models. In *Proceedings of the 3rd predictive methods conference*. Newport Beach, CA.

[4] Kushnir, T., Gopnik, A., Schulz, L. E., & Danks, D. (2003). Inferring hidden causes. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual meeting of the cognitive science society* (pp. 699-703). Boston: Cognitive Science Society.

[3] Danks, D. (2002). Learning the causal structure of overlapping variable sets. In S. Lange, K. Satoh, & C. H. Smith (Eds.), *Discovery science: Proceedings of the 5th international conference* (pp. 178-191). Berlin: Springer-Verlag.

[2] Danks, D., & Glymour, C. (2001). Linearity properties of Bayes nets with binary variables. In J. Breese & D. Koller (Eds.), *Uncertainty in artificial intelligence: Proceedings of the 17th conference (UAI-2001)* (pp. 98-104). San Francisco: Morgan Kaufmann.

[1] Wheeler, W., Danks, D., Ramsey, J., Scheines, R., Smith, J., & Thompson, A. (2001). Developing and deploying online courses with Jcourse. In *Proceedings of the association of the advancement of computing in education (AACE).*

Popular writings, whitepapers, and blogposts:

[18] Amico-Korby, D., Danks, D., & Harrell, M. (2023). Using verification to help social media users recognize experts. *Tech Policy Press*. https://techpolicy.press/using-verification-to-help-social-media-users-recognize-experts/

[17] Gentile, M. C., Danks, D., & Harrell, M. (2022). Case study: Does facial recognition tech enhance security? *Harvard Business Review*, Nov/Dec.

[16] Commentary on L. Kirfel & T. Gerstenberg target article for *The Brains Blog* (2022). https://philosophyofbrains.com/2022/09/20/cognitive-science-of-philosophy-symposium-causal-cognition.aspx

[15] Danks, D., & Harrell, M. (2022). *Ubiquitous Surveillance.* Giving Voice to Values case study A (UVA-OB-1403); case study B (UVA-OB-1404); and Teaching note (UVA-OB-1403TN).

[14] Data visualization video. 2021 Eradicate Hate Global Summit. https://www.youtube.com/watch?v=-Oi235nWc0U

[13] Persi Paoli. G., Vignard. K., Danks. D, & Meyer. P. (2020). *Modernizing arms control: Exploring responses to the use of AI in military decision-making*. Geneva, Switzerland: UNIDIR.

[12] Danks, D. (2020). Ubiquitous surveillance and the politics of refusal. In *Mirror with a memory: Photography, surveillance, and artificial intelligence.*

[11] Danks, D. (2020). How adversarial attacks could destabilize military AI systems. *IEEE Spectrum*, 26 February 2020.

[10] Danks, D., & Parker, J. (2019). *Ethical analysis of responses to synthetic and manipulated media.* Memo for Carnegie Endowment for International Peace.

[9] Danks, D., & Parker, J. (2019). *The un/ethical status of synthetic media.* Memo for Carnegie Endowment for International Peace.

[8] Major contributor for UNIDIR Observation Paper 9 (2019), *Algorithmic bias and the weaponization of increasingly autonomous technologies*. Geneva, Switzerland.

[7] Danks, D. (2019). Trust and values, bodies and AI. In catalog for *Paradox: Body in the age of AI* (art exhibit at the Miller Institute for Contemporary Art).

[6] Danks, D. (2018). AI & global governance: Supporting the ties that bind. UNU Centre for Policy Research, October 15, 2018. <https://cpr.unu.edu/ai-global-governance-supporting-the-ties-that-bind.html>

[5] Danks, D., & London, A. J. (2018). How driverless cars think. In "Forum." *Issues in Science and Technology, 34*(4).

[4] Taylor, S., Pickering, B., Boniface, M., Anderson, M., Danks, D., Følstad, A., Leese, M., Müller, V., Sorell, T., Winfield, A., & Woollard, F. (2018). Responsible AI: Key themes, concerns & recommendations for European research and innovation. DOI:10.5281/zenodo.1303253

[3] London, A. J., & Danks, D. (2017). Self-driving, but not self-regulating. *Pittsburgh Post-Gazette*, April 2, 2017. <http://www.post-gazette.com/opinion/Op-Ed/2017/04/02/Self-driving-but-not-self-regulating>

[2] Danks, D. (2016). Finding trust and understanding in autonomous technologies. *The Conversation*. December 30, 2016. <http://theconversation.com/finding-trust-and-understanding-in-autonomous-technologies-70245>

[1] Roff, H. M., Danks, D., & Danks, J. H. (2015). Fight ISIS by thinking inside the bot. *Slate*. October 21, 2015. <http://www.slate.com/articles/technology/future_tense/2015/10/using_chatbots_to_distract_isis_recruiters_on_social_media.html>

Book reviews and technical reports:

[R3] Danks, D. (2014). Review of *Perception, causation, & objectivity* (J. Roessler, H. Lerman, & N. Eilan, Eds.). *Mind, 123(490)*, 635-639.

[R2] Danks, D. (2005). Review of *Natural-born cyborgs: Minds, technologies, and the future of human intelligence* (A. Clark). *Philosophical Psychology, 18* (3), 383-387.

[R1] Danks, D. (2002). Review of *Graphical models: Foundations of neural computation* (M. I. Jordan & T. J. Sejnowski, Eds.). *Pattern Analysis and Applications, 5* (4), 401-402.

[T7] Hyttinen, A., Plis, S., Järvisalo, M., Eberhardt, F., & Danks, D. (2016). Causal discovery from subsampled time series data by constraint optimization. arXiv:1602.07970.

[T6] Davis, I., Kummerfeld, E., Danks, D., & Plis, S. (2015). Inferring observed structure for dynamic graphs with unobserved variables. Technical report CMU-PHIL-193. November 23, 2015.

[T5] Kummerfeld, E., & Danks, D. (2010). Online causal structure learning. Technical report CMU-PHIL-189. December 9, 2010.

[T4] Mayo-Wilson, C., Zollman, K., & Danks, D. (2010). Wisdom of the crowds vs. Groupthink: Learning in groups and in isolation. Technical report CMU-PHIL-188. November 30, 2010.

[T3] Chu, T., Danks, D., & Glymour, C. (2005). Data-driven methods for nonlinear Granger causality: Climate teleconnection mechanisms. Technical report CMU-PHIL-171. June 7, 2005.

[T2] Danks, D. (2004). Psychological theories of categorization as probabilistic models. Technical report CMU-PHIL-157. July 15, 2004.

[T1] Danks, D. (2003). Learning integrated structure from distributed databases with overlapping variables. Technical report CMU-PHIL-149. October 28, 2003.

Professional Presentations

224 professional presentations to date (61 peer-reviewed; 163 invited); full list available upon request

Fellowships, Awards, and Grants

External:

Multilevel Methodology to Develop Actionable RMAI Guidance 2024-25
Department of State
$161,062 (total costs); co-P.I. with N. Gallagher (Maryland)

In the Moment 2023-25
DARPA
$72,000 (UCSD total costs); sole P.I.

Algorithmic Bias in the Real World 2022-24

Office of Naval Research
$303,324 (total costs); sole P.I.

Sense-making Training for Information Analysts 2022-24
Office of Naval Research
$373,662 (total costs); co-P.I. with M. Harrell (UCSD)

Twiner Phase II 2022-23
Soar Technology
$303,867 (total costs); sole P.I.

CRCNS: Multimodal Dynamic Causal Learning for Neuroimaging 2021-25
NIMH
$286,205 (UCSD total costs); co-P.I. with S. Plis (Georgia State)

An Integrated Framework for Understanding Human-AI Hybrid Decision-Making 2021-22
Amazon Research Award (gift)
$103,000 (total costs); co-P.I. with S. Fazelpour & Z. Lipton (CMU)

Cognizant Center of Excellence on Content Moderation 2020-22
Cognizant Foundation
$750,000 (total costs); co-P.I. with K. Carley (CMU)

Center for Informed Democracy and Social Cybersecurity (IDeaS) 2019-25
John S. & James L. Knight Foundation
$5,000,000 (total costs); co-P.I. with K. Carley (CMU) & D. Sicker (CMU)

Autonomy and Moral Attribution 2019-21
Templeton World Charity Foundation
$234,000 (total costs); co-P.I. with T. Lombrozo (Princeton)

Misinformation Pipeline // Algorithmic Collusion 2019-20
Accenture Labs gift
$100,000 (direct costs); sole P.I.

Triple Variable Index 2019-20
Coulter Program, University of Pittsburgh
$100,000 (direct costs); co-P.I. with M. Schnetz, A. Mahajan, & M. Kaynar (UPMC)

Regulation of Defense and Security AI Technologies: Options Beyond Traditional Arms Control 2018-19
CIFAR workshop grant
$80,000 CAD (direct costs); co-P.I. with K. Vignard (UNIDIR) & P. Meyer (Simon Fraser)

CONTEXTS - Causal mOdeling for kNowledge Transfer, EXploration, and Temporal Simulation 2017-20
DARPA (subcontract to BAE Systems)
$548,918 (CMU total costs); sole CMU P.I.

Trust in an Age of Autonomous Technologies 2017-19
Andrew Carnegie Fellowship
$200,000 (direct costs); sole P.I.

Peer & Panel Review of Interdisciplinary Grant Proposals & Research Projects 2015-16
James S. McDonnell Foundation planning grant
$50,000 (direct costs); sole P.I.

Center for Causal Modeling and Discovery of Biomedical Knowledge from Big Data 2014-19
National Institutes of Health
$1,682,229 (CMU theory group total costs); co-I. with C. Glymour & P. Spirtes

Learning Causal Structure from Complex Time Series Data 2013-16
National Science Foundation
$217,497 (CMU total costs); co-P.I. with S. Plis (Mind Research Network)

Case Studies of Causal Discovery with Model Search 2012-13
National Science Foundation
$45,000 (total costs); co-P.I. with R. Scheines

Integrating Causal Cognition, Concepts, and Decision-making 2008-14
James S. McDonnell Foundation Scholar Award
$600,000 (direct costs); sole P.I.

Causal learning: Computational Learning Mechanisms and Cognitive Development 2005-10
James S. McDonnell Foundation Collaborative Initiative
$2.25 million (direct costs); one of 12 core members (Lead: A. Gopnik)

The Bayesian Network Lens 2002-03
James S. McDonnell Foundation grant
$49,615 (direct costs); sole P.I.

CMU Internal:

An Integrated Framework for Studying and Regulating Human-AI Hybrid Decision-Making Systems 2020-21
Block Center seed grant ($60,000; co-PI with Zack Lipton & Sina Fazelpour)

Trustworthy Transfer Learning: Determining the Limits of AI Robustness (DLAR) 2019-20
SEI LENS project ($350,000; co-I with PI Robert Stoddard)

Explanations, Trust, and AI 2018-20
Carnegie Bosch Institute Research Award ($200,000; co-I with PI Tae Wan Kim)

Integrated Causal Model for Software Cost Prediction & Control (SCOPE) 2017-20
SEI Line project ($2,000,000; co-I with PI Michael Konrad)

Innovating Air Force Jet Engine System Reliability Test using Machine Learning Integrated with Causal Modeling 2016-18
SEI Line project ($2,000,000; co-I with PI Robert Stoddard)

Why Does Software Cost So Much? Towards a Causal Model 2016-17
SEI LENS project ($500,000; co-I with PI Robert Stoddard)

Wimmer Faculty Fellowship (one of four) 2007

Travel grant from Berkman Faculty Development Fund to attend 2004
2004 International Congress of Psychology ($2,511)

Building Webs of Causal Knowledge (CMU Falk fellowship) 2003-05
$3,840 (direct costs) [sole P.I.]

Selected Professional Service [full lists for each category available upon request]

External:

One-Utah Responsible AI Initiative	External advisory board:	2024 -
AI Institute for Societal Decision Making (AI-SDM)	External advisory board:	2024 -
Centers for Communication Research (IDA) program management committee	Member:	2024 -
Responsible AI Academic Council (DoD CDAO)	Member:	2023 -
Computer Science and Telecommunications Board (NASEM)	Member:	2023 -

National Data Platform	Community Advisory Board member:	2023 -
Digital Safety Research Institute (DSRI) advisory board	Member:	2023 -
Philosophy of Science Association 2024 biennial meeting	Program chair:	2023 -
NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)	External advisory board:	2023 -
National Artificial Intelligence Advisory Committee (NAIAC)	Member:	2022 -
Ctr for Advancing Safety of Machine Intelligence (Northwestern)	Scientific Advisory Board:	2022 -
AI Community of Practice steering committee (U.S. government)	Founding member:	2022 -
Computing Community Consortium (CCC) Council, CRA	Member:	2022 -
	Executive Committee:	2022 -
Special Competitive Studies Project (SCSP), Society panel	Principal Advisor:	2021 -
Digital Society (Springer journal)	Scientific Advisory Board:	2021 -
Scientific Working Group Artificial Intelligence, Federal Bureau of Investigation	Member:	2021 -
Fellowships at Auschwitz for the Study of Professional Ethics	Academic Committee:	2019 -
Partnership to Advance Responsible Technology	Founding Board member:	2018 -
Salesforce Ethical & Responsible Use advisory council	External member:	2018 -
Philosophy of Cognitive Science, *Stanford Encyclopedia of Philosophy*	Co-editor:	2018 -
Minds & Machines	Editorial Board:	2010 -
Glushko Dissertation Prize selection committee	Member:	2018-23
	Chair:	2021-23
National Academies committee on Accelerating Behavioral Science Through Ontology Development and Use	Member:	2021-22
National Academies committee on Responsible Computing Research: Ethics and Governance of Computing Research and its Applications	Member:	2020-22
Technology Transformation Services (GSA) AI Portfolio	Expert advisor:	2020-22
Grefenstette Center for Ethics in Science, Tech., & Law (Duquesne)	Advisory Board:	2019-22
Pittsburgh Task Force on Public Algorithms	Member:	2019-22
Understanding Human Cognition program, James S. McDonnell Foundation	Advisory Board:	2010-22
IBM Watson AI XPRIZE competition	Lead/Presiding judge:	2017-21
National Security Commission on Artificial Intelligence	SGE for Ethics LOE:	2019-21
Frontiers in Psychology (Theoretical and Philosophical Psychology)	Associate editor:	2017-21
The Philosopher's Annual	Nominating Editor:	2008-18

<u>Internal to UCSD (excluding search committees):</u>

HDSI Faculty Council	Chair:	2023 -
Institute for Practical Ethics	Advisory Board:	2022 -
University of California AI Council	Member:	2021 -
Industry Liaison Partners (ILP) committee (HDSI)	Member:	2021-23
	Chair:	2022-23
Academic Senate	HDSI Representative:	2022-23

Internal to CMU:

Center for Informed Democracy and Social Cybersecurity (IDeaS)	Founding co-Director: 2019-21
Block Center for Technology & Society	Chief Ethicist: 2019-21
President's Task Force on Campus Climate	Co-chair: 2018-19
CMU Middle States reaccreditation process	Co-chair, Ethics & Integrity: 2016-18
Institutional Review Board [sabbatical in 2011-2012]	Member: 2004-14 Chair: 2009-14

Selected Workshop & Symposium co-organization (14 co-organized workshops total):

Workshop, "Models of morality, Morality of models" (2020)

Workshops, "Regulation of defense and security AI technologies: Options beyond traditional arms control" (2019/20 (two workshops); co-org)

2x CMU-K&L Gates Conferences on Ethics & Computational Technologies (2018, 2020; co-org)

Symposium at Philosophy of Science Association biennial meeting, "Bayesianism in cognitive science and neuroscience: An assessment" (2016; co-org)

Symposium at International Association on Computing & Philosophy meeting, "Automated and autonomous conflicts: AI, ethics, and the conduct of hostilities" (2016; co-org)

Special Session at AI & Mathematics, "Causal learning from complex data structures" (2012)

Workshop at NeurIPS, "Structured data and representations in probabilistic models for categorization" (2004; co-org)

Reviewing (multiple times for many):

57 distinct journals in philosophy, computer science, cognitive science, etc. (most multiple times)
Grant proposals from a variety of funders (including NSF, ESRC, ERC, NOW, ANR)
Multiple conferences (including 30 Program Committees)

Teaching Experience

University of California, San Diego

Undergraduate courses:

Ethics & Society II (PHIL)	[W-23]
Data Ethics (PHIL)	[S-22; W-23; S-24]
Living in a Digital World (PHIL)	[F-24]

Graduate seminars:

Ethics & AI (PHIL)	[F-21]

Graduate courses:

Data Science, Ethics, & Society (HDSI)	[W-22; S-22; S-23; F-24]
Data Fairness & Ethics (online) (HDSI)	[S-23]

Carnegie Mellon University

Graduate seminars:

Coherence	[F-17]
Computational Models of Cognition	[F-09]
Current Topics in Philosophy of Science	[F-07]
Ethics & Policy of AI	[S-19]
Ethics & Policy of Data Analytics	[S-21]

Graphical Models in Cognitive Science [S-06]
Normativity in Cognitive Psychology [S-04]
Philosophical Foundations (Core seminar) [12 times from S-08 to S-21]

Undergraduate courses:

AI & Ethics [F-19 micro]
AI, Society, & Humanity [F-18; F-19; F-20]
Cyberspace & Philosophy (Freshman seminar) [S-17]
Learning Media Principles [S-15; S-16; S-17; S-18]
The Nature of Reason [F-04; F-08]
Nietzsche [S-05; S-07; F-09; S-11; S-14; S-16; S-18]
Philosophy and Psychology [S-04; S-06; S-07; F-08]
Philosophy of Biology [S-05; S-08]
Philosophy of Mind [F-03; F-04; F-05]
Probability and Artificial Intelligence [F-07]
Thinking about Thinking (Freshman seminar) [S-13]

Colorado College

Philosophy of Biology [S-03]
Philosophy of Mind [F-02]

Thesis Advising and Committee Participation *[Full lists for each category available upon request]*

Primary Ph.D. (co-)advisor for 12 students
Primary M.S./M.A. (co-)advisor for 24 students
Primary undergraduate honor's thesis advisor for 3 students
Committee member/Second reader/External examiner for 41 students