Andrew Kyle Lampinen

Email
andrewlampinen@gmail.com

Website
https://lampinen.github.io

Education

Stanford University, Ph.D. Psychology (Cognitive), 2015-2020
- Center for Mind, Brain, Computation, and Technology Trainee.
- Minor in Computer Science.

UC Berkeley, B.A. Mathematics & Physics, 2010-2014

Research Positions

Staff Research Scientist, Google DeepMind, May 2024 - Present
Senior Research Scientist, Google DeepMind, March 2022 - Present
Research Scientist, DeepMind, October 2020 - February 2022
PhD Intern, DeepMind, May 2019 - September 2019
PhD Software Engineering Intern, Google Brain, June 2017 - September 2017
Associate Professional Staff I, Johns Hopkins University Applied Physics Laboratory, June 2014 - July 2015
Student Research Associate, Lawrence Berkeley National Laboratories, January - May 2012 & August - December 2012
Summer Research Intern, A*STAR Institute of High Performance Computing, Singapore, June - August 2012
Research Assistant, UC Davis Plant Sciences, June - August 2011

Honors

Cognitive Science Society Robert J. Glushko Dissertation Prize, 2021
Ric Weiland Graduate Fellowship in the Humanities and Sciences, 2018-2020
National Science Foundation Graduate Research Fellowship, 2015-2018
Percy Lionel Davis Award for Excellence in Scholarship in Mathematics, 2014
Berkeley Physics Olsen Scholar 2013-2014
Berkeley Letters & Science Dean's List 2012-2014
Berkeley Physics Undergraduate Research Scholar, Spring & Fall 2012

Selected Publications

Dan Friedman, **Andrew Lampinen**, Lucas Dixon, Danqi Chen, Asma Ghandeharioun (2024), "Interpretability illusions in the generalization of simplified models", *International Conference on Machine Learning*

Andrew K. Lampinen, Stephanie C. Y. Chan, Ishita Dasgupta, Andrew J. Nam, Jane X. Wang (2023), "Passive learning of active causal strategies in agents and language models", *Advances in Neural Information Processing Systems*

Lukas Muttenthaler, Lorenz Linhardt, Jonas Dippel, Robert A. Vandermeulen, Katherine Hermann, **Andrew K. Lampinen**, Simon Kornblith (2023), "Improving neural network representations using human similarity judgements", *Advances in Neural Information Processing Systems*

Wilka Carvalho, Andre Saraiva, Angelos Filos, **Andrew K. Lampinen**, Loic Matthey, Richard L. Lewis, Honglak Lee, Satinder Singh, Danilo J Rezende, Daniel Zoran (2023), "Combining behaviors with the successor features keyboard", *Advances in Neural Information Processing Systems*

Jerry Wei, Le Hou, **Andrew K. Lampinen**, Xiangning Chen, Da Huang, Yi Tay, Xinyun Chen, Yifeng Lu, Denny Zhou, Tengyu Ma, Quoc V. Le (2023), "Symbol tuning improves in-context learning in language models", *Proceedings of Empirical Methods in Natural Language Processing*

Aaditya K. Singh, David Ding, Andrew Saxe, Felix Hill, **Andrew K. Lampinen** (2023), "Know your audience: Specializing grounded language models with the game

of Dixit", *Proceedings of the European Chapter of the Association for Computational Linguistics*

Andrew K. Lampinen, Ishita Dasgupta, Stephanie C. Y. Chan, Kory Matthewson, Michael Henry Tessler, Antonia Creswell, James L. McClelland, Jane X. Wang, Felix Hill (2022), "Can language models learn from explanations in context?", *Findings of Empiricical Methods in Natural Language Processing*

Allison C. Tam, Neil. C. Rabinowitz, **Andrew K. Lampinen**, Nick A. Roy, Stephanie C. Y. Chan, DJ Strouse, Jane X. Wang, Andrea Banino, Felix Hill (2022), "Semantic exploration from language abstractions and pretrained representations", *Advances in Neural Information Processing Systems*

Stephanie C. Y. Chan, Adam Santoro, **Andrew K. Lampinen**, Jane X. Wang, Aaditya Singh, Pierre H. Richemond, James L. McClelland, Felix Hill (2022), "Data Distributional Properties Drive Emergent In-Context Learning in Transformers", *Advances in Neural Information Processing Systems*

Stephanie C. Y. Chan, Ishita Dasgputa, Junkyung Kim, Dharshan Kumaran, **Andrew K. Lampinen**, Felix Hill (2022), "Transformers generalize differently from information stored in context vs in weights", *Memory in Artificial & Natural Intelligence Workshop, NeurIPS 2022*

Andrew K. Lampinen, Nicholas A. Roy, Ishita Dasgupta, Stephanie C. Y. Chan, Allison C. Tam, James L. McClelland, Chen Yan, Adam Santoro, Neil C. Rabinowitz, Jane X. Wang, Felix Hill (2022), "Tell me why!—Explanations support learning of relational and causal structure", *International Conference on Machine Learning*

Stephanie C. Y. Chan*, **Andrew K. Lampinen***, Pierre H. Richemond*, Felix Hill* (2022), "Zipfian Environments for Reinforcement Learning", *Conference on Lifelong Learning Agents*, (*equal contribution)

Andrew K. Lampinen, Stephanie C. Y. Chan, Andrea Banino, Felix Hill (2021), "Towards mental time travel: a hierarchical memory for reinforcement learning agents", *Advances in Neural Information Processing Systems*

Andrew K. Lampinen, Stephanie C. Y. Chan, Adam Santoro, Felix Hill, (2021), "Publishing fast and slow: A path towards generalizability in psychology and AI", Commentary in *Behavioral and Brain Sciences*

Andrew K. Lampinen and James L. McClelland, (2020), "Transforming task representations to perform novel tasks", *Proceedings of the National Academy of Sciences*

Katherine L. Hermann* and **Andrew K. Lampinen***, (2020), "What shapes feature representations? Exploring datasets, architectures, and training", *Advances in Neural Information Processing Systems*, (*equal contribution)

James L. McClelland, Bruce L. McNaughton, and **Andrew K. Lampinen** (2020), "Integration of new information in memory: new insights from a complementary learning sytems perspective", *Proceedings of the Royal Society B*

Sébastien Racanière*, **Andrew K. Lampinen***, Adam Santoro, David P. Reichert, Vlad Firoiu, and Timothy P. Lillicrap, (2020), "Automated curricula through setter-solver interactions", *Proceedings of the 8th International Conference on Learning Representations*, (*equal contribution)

Felix Hill, **Andrew K. Lampinen**, Rosalia Schneider, Stephen Clark, Matthew Botvinick, James L. McClelland, and Adam Santoro (2020), "Environmental drivers of systematicity and generalisation in a situated agent", *Proceedings of the 8th International Conference on Learning Representations*

Andrew K. Lampinen and James L. McClelland, (2019), "Zero-shot task adaptation by homoiconic meta-mapping", *Learning Transferable Skills Workshop, NeurIPS*

Andrew K. Lampinen and Surya Ganguli, (2019), "An analytic theory of general-

ization dynamics and transfer learning in deep linear networks", *Proceedings of the 7th International Conference on Learning Representations*

Andrew K. Lampinen and James L. McClelland, (2018), "Different presentations of a mathematical concept can support learning in complementary ways", *Journal of Educational Psychology*

Robert X. D. Hawkins, Eric N. Smith, Carolyn Au, Juan Miguel Arias, Rhia Catapano, Eric Hermann, Martin Keil, **Andrew Lampinen**, Sarah Raposo, Jesse Reynolds, Shima Salehi, Justin Salloum, Jed Tan, and Michael C. Frank, (2018), "Improving the replicability of Psychological Science through pedagogy", *Advances in Methods and Practices in Psychological Science*

Steven S. Hansen, **Andrew K. Lampinen**, Gaurav Suri, and James L. McClelland, (2017), "Building on prior knowledge without building it in", *Commentary in Behavioral & Brain Sciences*

Andrew K. Lampinen, Shaw Hsu, and James L. McClelland, (2017), "Analogies emerge from learning dynamics in neural networks", *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*

Selected Preprints

Andrew K. Lampinen, Stephanie C. Y. Chan, Katherine Hermann (2024), "Learned feature representations are biased by complexity, learning order, position, and more", *arXiv*

Ilia Sucholutsky, Lukas Muttenthaler, Adrian Weller, Andi Peng, Andreea Bobu, Been Kim, Bradley C Love, Erin Grant, Jascha Achterberg, Joshua B Tenenbaum, Katherine M Collins, Katherine L Hermann, Kerem Oktar, Klaus Greff, Martin N Hebart, Nori Jacoby, Raja Marjieh, Robert Geirhos, Sherol Chen, Simon Kornblith, Sunayana Rane, Talia Konkle, Thomas P O'Connell, Thomas Unterthiner, **Andrew K. Lampinen***, Klaus-Robert Müller*, Mariya Toneva*, Thomas L Griffiths* (2023), "Getting aligned on representational alignment", *arXiv*, (*equal advising/senior authors)

Drew A Hudson, Daniel Zoran, Mateusz Malinowski, **Andrew K. Lampinen**, Andrew Jaegle, James L. McClelland, Loic Matthey, Felix Hill, Alexander Lerchner (2023), "SODA: Bottleneck Diffusion Models for Representation Learning", *arXiv*

Andrew K. Lampinen (2022), "Can language models handle recursively nested grammatical structures? A case study on comparing models and humans", *arXiv*

Adam Santoro*, **Andrew K. Lampinen***, Kory Mathewson, Timothy Lillicrap, David Raposo, (2021), "Symbolic Behaviour in Artificial Intelligence", *arXiv*, (*equal contribution)

Andrew K. Lampinen and James L. McClelland, (2017), "One-shot and few-shot learning of word embeddings", *arXiv*

Selected Invited Talks

"Symbolic behaviour in AI: some possible lessons for whales?", *Decoding Communication in Nonhuman Species Workshop, Simons Institute for the Theory of Computing*, June 2024

"What can be passively learned about causality?", *Understanding Higher-Level Intelligence Workshop, Simons Institute for the Theory of Computing*, June 2024 "Improving neural network representations by aligning with human knowledge", *Representational Alignment Workshop, ICLR 2024*, May 2024

"Comparing humans and language models: reasoning & grammar", *Cognition, Brain & Behavior Seminar, Harvard University*, February 2024

"A rose by any other representation: some questions on the relationship between representation & computation", *NeurIPS UniReps Workshop*, December 2023

"Comparing humans and language models: reasoning & grammar", *COLT Seminar, Universitat Pompeu Fabra*, November 2023

"Comparing humans and language models: reasoning & grammar", *UC San Diego Cognitive Science Seminar*, October 2023

"Comparing humans and language models: challenges & opportunities", *International Interdisciplinary Computational Cognitive Science Summer School*, September 2023

"Passive learning of active causal strategies", *Max Planck Institute for Biological Cybernetics*, June 2023

"Passive learning of active causal strategies", *Imperial College ICARL Seminar*, June 2023

"Comparing humans and language models: challenges & opportunities", *126th International Titisee Conference on NeuroAI*, March 2023

"Comparing humans and language models: reasoning & grammar", *Carnegie Mellon University BrAIn Seminar*, February 2023

"Language models show human-like content effects on reasoning", *London Machine Learning Meetup*, November 2022

"Augmenting reinforcement learning with language", *University of Edinburgh Computational Cognitive Science Seminar*, November 2022

"Compositionality (avoiding the question)", *Brown University*, October 2022

"Augmenting reinforcement learning with language", *University of Tokyo, Matsuo Lab*, August 2022

"Tell me why—Explanations improve learning of relational and causal structure", *Spotlight Presentation, International Conference on Machine Learning*, July 2022

"Tell me why—Explanations improve learning of relational and causal structure", *NYU Concepts & Categories Seminar*, February 2022

"A computational framework for learning and transforming task representations", *Cognitive Science Society Glushko Dissertation Prize Talk*, July 2021

"Task relationships, task transformations, and analogies", *Analogical Minds Seminar*, May 2021

"Multi-task learning, transfer, and abstraction", *Parallel Distributed Processing and the Emergence of an Understanding of Mind*, Princeton University, September 2018

"The Jabberwocky: One-shot and few-shot learning of word embeddings", *Meaning in Context Workshop*, Center for the Study of Language and Information, Stanford University, September 2017

Teaching Experience

Teaching Assistant, Stanford University Department of Psychology, 6 courses between Fall 2016 and Winter 2019

- Planned and taught discussion sections for undergraduate statistics & memory courses and graduate statistics & research methods courses.
- Gave lectures on reinforcement learning and wrote and graded homeworks for graduate course on Neural Network Models of Cognition.
- Held office hours.

Undergraduate Student Instructor, UC Berkeley Mathematics, Spring, Fall 2013, & Spring 2014

- Planned and taught discussion sections.
- Held office hours.
- Wrote and graded quizzes and midterms.

Teaching Assistant, UC Berkeley Early Academic Outreach Program, June-July 2013

- Held office hours.
- Substitute taught classes.

Other Work Experience

Statistics Consultant, Stanford University Department of Psychology, 2016-2017, 2019-2020

- Advised graduate students on technical aspects of data collection, analysis, and modeling.

Service

Action Editor:

- Transactions on Machine Learning Research

Area Chair:

- Neural Information Processing Systems

Reviewer:

- Artificial Intelligence
- Nature
- Nature Human Behavior
- Nature Machine Intelligence
- Nature Neuroscience
- Computational Linguistics
- Current Biology
- Neural Information Processing Systems
- International Conference on Learning Representations
- International Conference on Machine Learning
- Association for Computational Linguistics
- Cognitive Science Society
- Conference on the Mathematical Theory of Deep Neural Networks (DeepMath)
- Journal of Educational Psychology

Mentoring & other service:

- DeepMind Scholars Mentor
- Deep Indaba Mentor
- Team DE&I Lead, 2023
- Cientifico Latino Graduate Student Mentorship Initiative
- ICLR 2022 Co-Submitting Summer

Other Activities

Carillon: Carilloneur member of the Guild of Carilloneurs in North America (`www.gcna.org`).

Rock climbing: Bouldering, sport, and trad. Former routesetter at Stanford Climbing Wall, set problems for Collegiate Climbing Series events.