

Andrew M. Saxe

Center for Brain Science
Harvard University
Cambridge, MA, 02138 USA

Phone: 650-218-6701
Email: asaxe@fas.harvard.edu
URL: <http://www.people.fas.harvard.edu/~asaxe>

Nationality: American, British

Current position

Swartz Postdoctoral Fellow, Center for Brain Science, Harvard University

Research interests

The theory of deep learning and its applications to neuroscience and psychology.

Education

- 2015 PhD in Electrical Engineering, Stanford University (*June 2015*)
Advisors: James L. McClelland (primary), Surya Ganguli, Andrew Y. Ng, Christoph Schreiner
Thesis: “Deep linear neural networks: A theory of learning in the brain and mind”
- 2013-2014 Research Associate, Keck Center for Integrative Neuroscience, UCSF
- 2010 MS in Electrical Engineering, Stanford University
- 2008 BSE in Electrical Engineering, Princeton University, *summa cum laude*
Concentrations (minors): Robotics & Intelligent Systems; Applications of Computing; Applied and Computational Mathematics

Fellowships, honors & awards

- 2016 Robert J. Glushko Outstanding Doctoral Dissertations Prize, Cognitive Science Society
- 2013-15 Center for Mind, Brain, and Computation Traineeship
- 2013 Artificial Intelligence Journal Travel Award, CogSci2013
- 2010-13 National Defense Science and Engineering Graduate (NDSEG) Fellowship
- 2010 NSF Graduate Research Fellowship Honorable Mention
- 2008-10 Stanford Graduate Fellowship, Stanford University
- 2008 Hertz Fellowship Finalist
- 2008 Lore von Jaskowsky Memorial Prize for Contributions to Research, Princeton University
- 2008 G. David Forney Jr. Prize in Signals & Systems, Princeton University
- 2008 Phi Beta Kappa, Princeton University
- 2007-8 Barry M. Goldwater Scholarship
- 2007 Schapiro Prize for Academic Excellence, Princeton University

Refereed publications

- Earle, A.C., A.M. Saxe, and B. Rosman (2017). “Hierarchical Subtask Discovery With Non-Negative Matrix Factorization”. In: *Workshop on Lifelong Learning: A Reinforcement Learning Approach at ICML*.
- Musslick, S. et al. (2017). “Multitasking Capability Versus Learning Efficiency in Neural Network Architectures”. In: *Proceedings of the 39th annual meeting of the Cognitive Science Society*, pp. 829–834.
- Saxe, A.M., A.C. Earle, and B. Rosman (2017). “Hierarchy Through Composition with Multitask LMDPs”. In: *Proceedings of the 34th International Conference on Machine Learning*. Sydney, Australia.
- McClelland, J.L., Z. Sadeghi, and A.M. Saxe (2016). “A Critique of Pure Hierarchy: Uncovering Cross-Cutting Structure in a Natural Dataset”. In: *Neurocomputational Models of Cognitive Development and Processing*, pp. 51–68.
- Tsai*, C.Y., A. Saxe*, and D. Cox (2016). “Tensor Switching Networks”. In: *Advances in Neural Information Processing Systems 29*. *Equal contributions.
- Goodfellow, I.J., O. Vinyals, and A.M. Saxe (2015). “Qualitatively Characterizing Neural Network Optimization Problems”. In: *the International Conference on Learning Representations*. Oral presentation. San Diego, CA.
- Saxe, A.M., J.L. McClelland, and S. Ganguli (2014). “Exact solutions to the nonlinear dynamics of learning in deep linear neural networks”. In: *the International Conference on Learning Representations*. Ed. by Y. Bengio and Y. LeCun. Oral presentation. Banff, Canada.
- Saxe, A.M., J.L. McClelland, and S. Ganguli (2013b). “Dynamics of learning in deep linear neural networks”. In: *NIPS Workshop on Deep Learning*.
- Saxe, A.M., J.L. McClelland, and S. Ganguli (2013c). “Learning hierarchical category structure in deep neural networks”. In: *Proceedings of the 35th annual meeting of the Cognitive Science Society*. Ed. by M. Knauff et al. Oral presentation. Austin, TX: Cognitive Science Society, pp. 1271–1276.
- Balci, F. et al. (2011). “Acquisition of decision making criteria: reward rate ultimately beats accuracy”. In: *Attention, Perception, & Psychophysics* 73.2, pp. 640–657.
- Saxe, A.M. et al. (2011). “On Random Weights and Unsupervised Feature Learning”. In: *Proceedings of the 28th International Conference on Machine Learning*.
- Saxe, A. et al. (2011). “Unsupervised learning models of primary cortical receptive fields and receptive field plasticity”. In: *Advances in Neural Information Processing Systems 25*.
- Saxe, A.M. et al. (2010). “On Random Weights and Unsupervised Feature Learning”. In: *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*.
- Baldassano, C.A. et al. (2009). “Kratos: Princeton University’s entry in the 2008 Intelligent Ground Vehicle Competition”. In: *Proceedings of SPIE*.
- Goodfellow, I.J., Q.V. Le, et al. (2009). “Measuring Invariances in Deep Networks”. In: *Advances in Neural Information Processing Systems 24*. Ed. by Y. Bengio and D. Schuurmans.
- Atreya, A.R. et al. (2006). “Prospect Eleven: Princeton University’s entry in the 2005 DARPA Grand Challenge”. In: *Journal of Field Robotics* 23.9, pp. 745–753.

Preprints

- Advani*, M. and A.M. Saxe* (2017). “High-dimensional dynamics of generalization error in neural networks”. In: *arXiv*. *Equal contributions.

Refereed conference abstracts

- Baldassano*, C. and A.M. Saxe* (2016). “A theory of learning dynamics in perceptual decision-making”. In: *the Computational and Systems Neuroscience Conference*. *Equal contributions. Salt Lake City.
- Saxe, A.M. and K. Norman (2016). “Optimal storage capacity associative memories exhibit retrieval-induced forgetting”. In: *the Computational and Systems Neuroscience Conference*. Salt Lake City.
- Lee, R. and A.M. Saxe (2015). “The Effect of Pooling in a Deep Learning Model of Perceptual Learning”. In: *the Computational and Systems Neuroscience Conference*. Salt Lake City.
- Saxe, A.M. (2015). “A deep learning theory of perceptual learning dynamics”. In: *the Computational and Systems Neuroscience Conference*. Salt Lake City.
- Saxe, A.M., J.L. McClelland, and S. Ganguli (2013a). “A Mathematical Theory of Semantic Development”. In: *the Computational and Systems Neuroscience Conference (COSYNE)*. Salt Lake City.
- Saxe, A.M., M. Bhand, et al. (2011). “Modeling Cortical Representational Plasticity With Unsupervised Feature Learning”. In: *the Computational and Systems Neuroscience Conference (COSYNE)*.

Invited presentations

- 2017 June Temporal Dynamics of Learning Seminar, UCSD, San Diego, CA
- 2016 Sep Google DeepMind, London
- 2016 Aug 15th Neural Computation and Psychology Workshop, Philadelphia
- 2016 Jul Google Research, Cambridge, MA
- 2016 Jun Deep Learning Workshop, Center for Brains, Minds, and Machines, MIT
- 2016 Feb Redwood Center for Theoretical Neuroscience, UC Berkeley
- 2016 Feb Apple, Cupertino, CA
- 2015 Dec Brains, Minds, and Machines Symposium, NIPS, Montreal

Other presentations

- Saxe, A.M. (2016). “Inferring actions, intentions, and causal relations in a neural network”. In: *Proceedings of the 38th annual meeting of the Cognitive Science Society*. Philadelphia.
- Lee, R., A.M. Saxe, and J. McClelland (2014). *Modeling Perceptual Learning with Deep Networks*. Quebec City.
- Saxe, A.M. (2014). “Multitask Model-free Reinforcement Learning”. In: *the 36th annual meeting of the Cognitive Science Society*. Quebec City.

Teaching

- 2017 Course Designer, Introductory Computational Neuroscience, Harvard University
- 2017 Distinction in Teaching Award (MCB131), Harvard University
- 2017 Head Teaching Fellow, MCB131: Computational Neuroscience, Harvard University
- 2016 Coadvisor for doctoral candidate, University of the Witwatersrand, SA
- 2013-15 Mentor for Independent Study courses involving undergraduate and master students, Stanford University
- 2014 Guest Lecturer, PSYCH209: Neural network and deep learning models for cognition and cognitive neuroscience, Stanford University
- 2010 Teaching Assistant, CS294A: Research projects in Artificial Intelligence, Stanford University
- 2009 Teaching Assistant, CS229: Machine Learning, Stanford University

Service activities

JOURNAL REVIEWER

Nature Communications

Proceedings of the National Academy of Sciences (PNAS)

Journal of Machine Learning Research (JMLR)

PLOS ONE

Neural Computation

IEEE Transactions on Neural Networks and Learning Systems (IEEE-TNNLS)

IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE-TPAMI)

IEEE Transactions on Knowledge and Data Engineering (IEEE-TKDE)

CONFERENCE REVIEWER

International Conference on Machine Learning (ICML)

Advances in Neural Information Processing Systems (NIPS) (Reviewer Award, 2013)

International Conference on Learning Representations (ICLR) (Reviewer Award, 2017)

International Conference on Artificial Intelligence and Statistics (AISTATS)

Cognitive Science Society Annual Meeting (CogSci)

WORKSHOP ORGANIZER

2016

CogSci 2016 Tutorial Workshop on Contemporary Deep Neural Network Models, Philadelphia

2014

CogSci 2014 Workshop on Deep Learning and the Brain, Quebec City, Canada