Andrew M. Saxe

Department of Experimental Psychology
University of Oxford

Phone: +44 07541866812
Email: andrew.saxe@psy.ox.ac.uk
Nationality: American, British

Current position

Sir Henry Dale Fellow, Department of Experimental Psychology, Oxford University

Research interests

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

Past positions and academic training

2018-19 *Postdoctoral Research Associate*, Department of Experimental Psychology, Oxford University
2015-18 *Swartz Postdoctoral Fellow*, Center for Brain Science, Harvard University

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

2019 Wellcome-Beit Prize, Wellcome Trust
2019-24 Sir Henry Dale Fellowship, Wellcome Trust & Royal Society
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

Refereed publications

Goldt, S. et al. (2019). “Generalisation dynamics of online learning in over-parameterised neural networks”. In: *ICML Workshop on Theoretical Physics for Deep Learning Theory.*

Goldt, Sebastian et al. (2019). “Dynamics of stochastic gradient descent for two-layer neural networks in the teacher-student setup”. In: *NeurIPS*. Oral presentation.

Richards, Blake A. et al. (2019). “A deep learning framework for neuroscience”. In: *Nature Neuroscience* 22.11, pp. 1761–1770.

Saxe, Andrew M., James L. McClelland, and Surya Ganguli (2019). “A mathematical theory of semantic development in deep neural networks”. In: *Proceedings of the National Academy of Sciences* 116.23, pp. 11537–11546.

Earle, A.C., A.M. Saxe, and B. Rosman (2018). “Hierarchical Subtask Discovery with Non-Negative Matrix Factorization”. In: *International Conference on Learning Representations.* Ed. by Y. Bengio and Y. LeCun. Vancouver, Canada.

Nye, M. and A. Saxe (2018). “Are Efficient Deep Representations Learnable?” In: *Workshop Track at the International Conference on Learning Representations.* Ed. by Y. Bengio and Y. LeCun. Vancouver, Canada.

Saxe, A.M., Y. Bansal, et al. (2018). “On the Information Bottleneck Theory of Deep Learning”. In: *International Conference on Learning Representations.* Ed. by Y. Bengio and Y. LeCun. Vancouver, Canada.

Zhang, Y. et al. (2018). “Energy-entropy competition and the effectiveness of stochastic gradient descent in machine learning”. In: *Molecular Physics*, pp. 1–10.

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: *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: *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: *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: *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: *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. 640–57.

Saxe, A.M., P.W. Koh, 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., P.W. Koh, 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

Bansal, Y. et al. (2018). “Minnorm training: an algorithm for training over-parameterized deep neural networks”. In: *arXiv.*

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

Refereed conference abstracts

Masís, J., A.M. Saxe, and D.D. Cox (2018). “Rats optimize reward rate and learning speed in a 2-AFC task”. In: *Computational and Systems Neuroscience Conference.* Denver.

Saxe*, A.M. and M. Advani* (2018). “A theory of memory replay and generalization performance in neural networks”. In: *Computational and Systems Neuroscience Conference.* *Equal contributions. Denver.

Baldassano*, C. and A.M. Saxe* (2016). “A theory of learning dynamics in perceptual decision-making”. In: *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: *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: *Computational and Systems Neuroscience Conference.* Salt Lake City.

Saxe, A.M. (2015). “A deep learning theory of perceptual learning dynamics”. In: *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: *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: *Computational and Systems Neuroscience Conference (COSYNE).*

Invited presentations

2018 Nov	Statistical Physics Seminar, ENS, Paris
2018 Sep	PDP Symposium, Princeton
2018 Sep	Computation and Theory Seminar, Janelia
2018 Feb	Symposium on the Mathematical Theory of Deep Neural Networks, Princeton
2017 Dec	Oxford Neurotheory Forum, Oxford
2017 Jun	Temporal Dynamics of Learning Seminar, UCSD
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: *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: *Annual meeting of the Cognitive Science Society*. Quebec City.

Teaching

2018	Distinction in Teaching Award (NEURO120), Harvard University
2017	Course Designer, Introductory Computational Neuroscience (NEURO120), 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, Undergraduate Honors Thesis, 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 & 2017)
International Conference on Learning Representations (ICLR) (Reviewer Award, 2017)
International Conference on Artificial Intelligence and Statistics (AISTATS)
Cognitive Science Society Annual Meeting (CogSci)

Conference Organizer

2019 Conference on the Mathematical Theory of Deep Neural Networks, New York

2019 Conference on Deep Learning and the Brain, Jerusalem, Israel

Workshop Organizer

2019 Cosyne 2019 Workshop on continual learning in biological and artificial neural networks

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

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