[1] P. Abbeel and J. Schulman ,“Deep Reinforcement Learning through Policy Optimization,”  *Tutorial at NIPS 2016.*  
[2] K. Arulkumaran, N. Dilokthanakul, M. Shanahan, and A.A. Bharath ,“Classifying Options for Deep Reinforcement Learning,” *IJCAI Workshop on Deep Reinforcement Learning: Frontiers and Challenges*, 2016  
[3] P.L. Bacon, J. Harb, and D. Precup ,“The Option-Critic Architecture,” *AAAI*, 2017  
[4] J.A. Bagnell and J. Schneider ,“Covariant Policy Search,” *IJCAI*, 2003  
[5] D. Bahdanau, P. Brakel, K. Xu, A. Goyal, R. Lowe, J. Pineau, A. Courville, and Y. Bengio ,“An Actor-Critic Algorithm for Sequence Prediction,” *ICLR*, 2017  
[6] L.C. Baird~ I I I ,“Advantage Updating,”  *Technical report, DTIC, 1993.*  
[7] N. Baram, O. Anschel, and S. Mannor ,“Model-Based Adversarial Imitation Learning,” *NIPS Workshop on Deep Reinforcement Learning*, 2016  
[8] A.G. Barto, R.S. Sutton, and C.W. Anderson ,“Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems,” *IEEE Trans. on Systems, Man, and Cybernetics*, iss.5, pp.834-846, 1983  
[9] M.G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling ,“The Arcade Learning Environment: An Evaluation Platform for General Agents,” *IJCAI*, 2015  
[10] M.G. Bellemare, S. Srinivasan, G. Ostrovski, T. Schaul, D. Saxton, and R. Munos ,“Unifying Count-Based Exploration and Intrinsic Motivation,” *NIPS*, 2016  
[11] M.G. Bellemare, W. Dabney, and R. Munos ,“A Distributional Perspective on Reinforcement Learning,” *ICML*, 2017  
[12] R. Bellman ,“On the Theory of Dynamic Programming,” *PNAS*, vol.38, iss.8, pp.716-719, 1952  
[13] Y. Bengio, J. Louradour, R. Collobert, and J. Weston ,“Curriculum Learning,” *ICML*, 2009  
[14] Y. Bengio, A. Courville, and P. Vincent ,“Representation Learning: A Review and New Perspectives,” *IEEE Trans. on Pattern Analysis and Machine intelligence*, vol.35, iss.8, pp.1798-1828, 2013  
[15] D.P. Bertsekas ,“Dynamic Programming and Suboptimal Control: A Survey from ADP to MPC,” *European Journal of Control*, vol.11, iss.4-5, pp. 310-334, 2005  
[16] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba ,“OpenAI Gym,” *arXiv:1606.01540*, 2016  
[17] L. Busoniu, R. Babuska, and B.D. Schutter ,“A Comprehensive survey of Multiagent Reinforcement Learning,” *IEEE Trans. on Systems, Man, And Cybernetics*, 2008  
[18] M. Campbell, A.J. Hoane, and F. Hsu ,“Deep Blue,” *Artificial intelligence*, vol.134, iss.1-2, pp.57-83, 2002  
[19] R. Caruana ,“Multitask Learning,” *Machine Learning*, vol.28, iss.1, pp.41-75, 1997  
[20] S. Chiappa, S. Racaniere, D. Wierstra, and S. Mohamed ,“Recurrent Environment Simulators,” *ICLR*, 2017  
[21] P. Christiano, Z. Shah, I. Mordatch, J. Schneider, T. Blackwell, J. Tobin, P. Abbeel, and W. Zaremba ,“Transfer from Simulation to Real World through Learning Deep Inverse Dynamics Model,” *arXiv:1610.03518*, 2016  
[22] G. Cuccu, M. Luciw, J. Schmidhuber, and F. Gomez ,“Intrinsically Motivated Neuroevolution for Vision-Based Reinforcement Learning,” *ICDL*, 2011  
[23] P. Dayan ,“Improving Generalization for Temporal Difference Learning: The Successor Representation,” *Neural Computation*, vol.5, iss.4, pp.613-624, 1993  
[24] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, Q.V. Le, and et~al ,“Large Scale Distributed Deep Networks,” *NIPS*, 2012  
[25] M.P. Deisenroth, G. Neumann, and J. Peters ,“A Survey on Policy Search for Robotics,” *Foundations and Trends{\textregistered*, vol.2, iss.1--2, 2013  
[26] M. Denil, P. Agrawal, T.D. Kulkarni, T. Erez, P. Battaglia, and N. Freitas ,“Learning to Perform Physics Experiments via Deep Reinforcement Learning,” *ICLR*, 2017  
[27] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel ,“Benchmarking Deep Reinforcement Learning for Continuous Control,” *ICML*, 2016  
[28] Y. Duan, J. Schulman, X. Chen, P.L. Bartlett, I. Sutskever, and P. Abbeel ,“RL$^2$: Fast Reinforcement Learning via Slow Reinforcement Learning,” *NIPS Workshop on Deep Reinforcement Learning*, 2016  
[29] G.D. Arnold, R. Evans, H. Hasselt, P. Sunehag, T. Lillicrap, J. Hunt, T. Mann, T. Weber, T. Degris, and B. Coppin ,“Deep Reinforcement Learning in Large Discrete Action Spaces,” *arXiv:1512.07679*, 2015  
[30] D. Ferrucci, E. Brown, J.C. Carroll, J. Fan, D. Gondek, A.A. Kalyanpur, A. Lally, J.W. Murdock, E. Nyberg, J. Prager, and et~al ,“Building Watson: An Overview of the DeepQA Project,” *AI Magazine*, vol.31, iss.3, pp.59-79, 2010  
[31] C. Finn, X.Y. Tan, Y. Duan, T. Darrell, S. Levine, and P. Abbeel ,“Deep Spatial Autoencoders for Visuomotor Learning,” *ICRA*, 2016  
[32] J. Foerster, Y.M. Assael, N. Freitas, and S. Whiteson ,“Learning to Communicate with Deep Multi-Agent Reinforcement Learning,” *NIPS*, 2016  
[33] M.C. Fu ,“Gradient Estimation,” *Handbooks in Operations Research and Management Science*, vol.13:, pp.575-616, 2006  
[34] M. Garnelo, K. Arulkumaran, and M. Shanahan ,“Towards Deep Symbolic Reinforcement Learning,” *NIPS Workshop on Deep Reinforcement Learning*, 2016  
[35] P.W. Glynn ,“Likelihood Ratio Gradient Estimation for Stochastic Systems,” *Communications of the ACM*, vol.33, iss.10, pp.75-84, 1990  
[36] F. Gomez and J. Schmidhuber ,“Evolving Modular Fast-Weight Networks for Control,” *ICANN*, 2005  
[37] I. Goodfellow, J.P. Abadie, M. Mirza, B. Xu, D.W. Farley, S. Ozair, A. Courville, and Y. Bengio ,“Generative Adversarial Nets,” *NIPS*, 2014  
[38] A. Gruslys, M.G. Azar, M.G. Bellemare, and R. Munos ,“The Reactor: A Sample-Efficient Actor-Critic Architecture,” *arXiv:1704.04651*, 2017  
[39] S. Gu, T. Lillicrap, I. Sutskever, and S. Levine ,“Continuous Deep Q-Learning with Model-Based Acceleration,” *ICLR*, 2016  
[40] S. Gu, T. Lillicrap, Z. Ghahramani, R.E. Turner, and S. Levine ,“Q-Prop: Sample-Efficient Policy Gradient with an Off-Policy Critic,” *ICLR*, 2017  
[41] M.E. Harmon and L.C. Baird~ I I I ,“Multi-Player Residual Advantage Learning with General Function Approximation,”  *Technical report, DTIC, 1996.*  
[42] M. Hausknecht and P. Stone ,“Deep Recurrent Q-Learning for Partially Observable MDPs,” *AAAI Fall Symposium Series*, 2015  
[43] N. Heess, J.J. Hunt, T.P. Lillicrap, and D. Silver ,“Memory-Based Control with Recurrent Neural Networks,” *NIPS Workshop on Deep Reinforcement Learning*, 2015  
[44] N. Heess, G. Wayne, D. Silver, T. Lillicrap, T. Erez, and Y. Tassa ,“Learning Continuous Control Policies by Stochastic Value Gradients,” *NIPS*, 2015  
[45] N. Heess, S. Sriram, J. Lemmon, J. Merel, G. Wayne, Y. Tassa, T. Erez, Z. Wang, A. Eslami, M. Riedmiller, and et~al ,“Emergence of Locomotion Behaviours in Rich Environments,” *arXiv:1707.02286*, 2017  
[46] J. Heinrich and D. Silver ,“Deep Reinforcement Learning from Self-Play in Imperfect-Information Games,”  *2016.*  
[47] T. Hester, M. Vecerik, O. Pietquin, M. Lanctot, T. Schaul, B. Piot, A. Sendonaris, G.D. Arnold, I. Osband, J. Agapiou, and et~al ,“Learning from Demonstrations for Real World Reinforcement Learning,” *arXiv:1704.03732*, 2017  
[48] G. Hinton, O. Vinyals, and J. Dean ,“Distilling the Knowledge in a Neural Network,”  *2014.*  
[49] J. Ho and S. Ermon ,“Generative Adversarial Imitation Learning,” *NIPS*, 2016  
[50] R. Houthooft, X. Chen, Y. Duan, J. Schulman, F.D. Turck, and P. Abbeel ,“VIME: Variational Information Maximizing Exploration,” *NIPS*, 2016  
[51] A. Hussein, M.M. Gaber, and E. Elyan ,“Deep Active Learning for Autonomous Navigation,” *EANN*, 2016  
[52] M. Jaderberg, V. Mnih, W.M. Czarnecki, T. Schaul, J.Z. Leibo, D. Silver, and K. Kavukcuoglu ,“Reinforcement Learning with Unsupervised Auxiliary Tasks,” *ICLR*, 2017  
[53] M. Johnson, K. Hofmann, T. Hutton, and D. Bignell ,“The Malmo Platform for Artificial Intelligence Experimentation,” *IJCAI*, 2016  
[54] L.P. Kaelbling, M.L. Littman, and A.R. Cassandra ,“Planning and Acting in Partially Observable Stochastic Domains,” *Artificial intelligence*, vol.101, iss.1, pp.99-134, 1998  
[55] S.M. Kakade ,“A Natural Policy Gradient,” *NIPS*, 2002  
[56] K. Kansky, T. Silver, D.A. Mély, M. Eldawy, M.L. Gredilla, X. Lou, N. Dorfman, S. Sidor, S. Phoenix, and D. George ,“Schema Networks: Zero-Shot Transfer with a Generative Causal Model of Intuitive Physics,” *ICML*, 2017  
[57] H.J. Kappen ,“Path Integrals and Symmetry Breaking for Optimal Control Theory,” *JSTAT*, vol.2005, iss.11, 2005  
[58] D.P. Kingma and M. Welling ,“Auto-Encoding Variational Bayes,” *ICLR*, 2014  
[59] N. Kohl and P. Stone ,“Policy Gradient Reinforcement Learning for Fast Quadrupedal Locomotion,” *ICRA*, 2004  
[60] V.R. Konda and J.N. Tsitsiklis ,“On Actor-Critic Algorithms,” *SICON*, vol.42, iss.4, pp.1143-1166, 2003  
[61] J. Koutník, G. Cuccu, J. Schmidhuber, and F. Gomez ,“Evolving Large-Scale Neural Networks for Vision-Based Reinforcement Learning,” *GECCO*, 2013  
[62] T.D. Kulkarni, K. Narasimhan, A. Saeedi, and J. Tenenbaum ,“Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation,” *NIPS*, 2016  
[63] T.D. Kulkarni, A. Saeedi, S. Gautam, and S.J. Gershman ,“Deep Successor Reinforcement Learning,” *NIPS Workshop on Deep Reinforcement Learning*, 2016  
[64] T.L. Lai and H. Robbins ,“Asymptotically Efficient Adaptive Allocation Rules,” *Advances in Applied Mathematics*, vol.6, iss.1, pp. 4-22, 1985  
[65] B.M. Lake, T.D. Ullman, J.B. Tenenbaum, and S.J. Gershman ,“Building Machines That Learn and Think Like People,” *The Behavioral and Brain Sciences*, 2016  
[66] S. Lange, M. Riedmiller, and A. Voigtlander ,“Autonomous Reinforcement Learning on Raw Visual Input Data in a Real World Application,” *IJCNN*, 2012  
[67] Y.L. Cun, Y. Bengio, and G. Hinton ,“Deep Learning,” *Nature*, vol.521, iss.7553, pp.436-444, 2015  
[68] J.Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, and T. Graepel ,“Multi-Agent Reinforcement Learning in Sequential Social Dilemmas,” *AAMAS*, 2017  
[69] S. Levine and P. Abbeel ,“Learning Neural Network Policies with Guided Policy Search under Unknown Dynamics,” *NIPS*, 2014  
[70] S. Levine and V. Koltun ,“Guided Policy Search,” *ICLR*, 2013  
[71] S. Levine, C. Finn, T. Darrell, and P. Abbeel ,“End-to-End Training of Deep Visuomotor Policies,” *JMLR*, vol.17, iss.39, pp.1-40, 2016  
[72] S. Levine, P. Pastor, A. Krizhevsky, and D. Quillen ,“Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection,” *ISER*, 2016  
[73] K. Li and J. Malik ,“Learning to Optimize,”  *2017.*  
[74] X. Li, L. Li, J. Gao, X. He, J. Chen, L. Deng, and J. He ,“Recurrent Reinforcement Learning: A Hybrid Approach,” *arXiv:1509.03044*, 2015  
[75] Y. Li ,“Deep Reinforcement Learning: An Overview,” *arXiv:1701.07274*, 2017  
[76] T.P. Lillicrap, J.J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra ,“Continuous Control with Deep Reinforcement Learning,” *ICLR*, 2016  
[77] L.J. Lin ,“Self-Improving Reactive Agents Based on Reinforcement Learning, Planning and Teaching,” *Machine Learning*, vol.8, iss.3--4, pp.293-321, 1992  
[78] L. Metz, J. Ibarz, N. Jaitly, and J. Davidson ,“Discrete Sequential Prediction of Continuous Actions for Deep RL,” *arXiv:1705.05035*, 2017  
[79] P. Mirowski, R. Pascanu, F. Viola, H. Soyer, A. Ballard, A. Banino, M. Denil, R. Goroshin, L. Sifre, K. Kavukcuoglu, and et~al ,“Learning to Navigate in Complex Environments,” *ICLR*, 2017  
[80] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu ,“Recurrent Models of Visual Attention,” *NIPS*, 2014  
[81] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, and et~al ,“Human-Level Control through Deep Reinforcement Learning,” *Nature*, vol.518, iss.7540, pp.529-533, 2015  
[82] V. Mnih, A.P. Badia, M. Mirza, A. Graves, T.P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu ,“Asynchronous Methods for Deep Reinforcement Learning,” *ICLR*, 2016  
[83] S. Mohamed and D.J. Rezende ,“Variational Information Maximisation for Intrinsically Motivated Reinforcement Learning,” *NIPS*, 2015  
[84] A.W. Moore ,“Efficient Memory-Based Learning for Robot Control,”  *Technical report, University of Cambridge, Computer Laboratory, 1990.*  
[85] R. Munos, T. Stepleton, A. Harutyunyan, and M.G. Bellemare ,“Safe and Efficient Off-Policy Reinforcement Learning,” *NIPS*, 2016  
[86] O. Nachum, M. Norouzi, K. Xu, and D. Schuurmans ,“Bridging the Gap Between Value and Policy Based Reinforcement Learning,” *arXiv:1702.08892*, 2017  
[87] A. Nagabandi, G. Kahn, R.S. Fearing, and S. Levine ,“Neural Network Dynamics for Model-Based Deep Reinforcement Learning with Model-Free Fine-Tuning,” *arXiv:1708.02596*, 2017  
[88] A. Nair, P. Srinivasan, S. Blackwell, C. Alcicek, R. Fearon, A.D. Maria, V. Panneershelvam, M. Suleyman, C. Beattie, S. Petersen, and et~al ,“Massively Parallel Methods for Deep Reinforcement Learning,” *ICML Workshop on Deep Learning*, 2015  
[89] A.Y. Ng and S.J. Russell ,“Algorithms for Inverse Reinforcement Learning,” *ICML*, 2000  
[90] A.Y. Ng, A. Coates, M. Diel, V. Ganapathi, J. Schulte, B. Tse, E. Berger, and E. Liang ,“Autonomous Inverted Helicopter Flight via Reinforcement Learning,” *Experimental Robotics*, 2006  
[91] B.O. Donoghue, R. Munos, K. Kavukcuoglu, and V. Mnih ,“PGQ: Combining Policy Gradient and Q-Learning,” *ICLR*, 2017  
[92] J. Oh, X. Guo, H. Lee, R.L. Lewis, and S. Singh ,“Action-Conditional Video Prediction using Deep Networks in Atari Games,” *NIPS*, 2015  
[93] J. Oh, V. Chockalingam, S. Singh, and H. Lee ,“Control of Memory, Active Perception, and Action in Minecraft,” *ICLR*, 2016  
[94] I. Osband, C. Blundell, A. Pritzel, and B.V. Roy ,“Deep Exploration via Bootstrapped DQN,” *NIPS*, 2016  
[95] E. Parisotto and R. Salakhutdinov ,“Neural Map: Structured Memory for Deep Reinforcement Learning,” *arXiv:1702.08360*, 2017  
[96] E. Parisotto, J.L. Ba, and R. Salakhutdinov ,“Actor-Mimic: Deep Multitask and Transfer Reinforcement Learning,” *ICLR*, 2016  
[97] D. Pathak, P. Agrawal, A.A. Efros, and T. Darrell ,“Curiosity-Driven Exploration by Self-supervised Prediction,” *ICML*, 2017  
[98] P. Peng, Y. Wen, Y. Yang, Q. Yuan, Z. Tang, H. Long, and J. Wang ,“Multiagent Bidirectionally-Coordinated Nets: Emergence of Human-level Coordination in Learning to Play StarCraft Combat Games,” *arXiv:1703.10069*, 2017  
[99] D.A. Pomerleau ,“ALVINN, an Autonomous Land Vehicle in a Neural Network,”  *Technical report, Carnegie Mellon University, Computer Science Department, 1989.*  
[100] A. Pritzel, B. Uria, S. Srinivasan, A. Puigdomènech, O. Vinyals, D. Hassabis, D. Wierstra, and C. Blundell ,“Neural Episodic Control,” *ICML*, 2017  
[101] M.A. Ranzato, S. Chopra, M. Auli, and W. Zaremba ,“Sequence Level Training with Recurrent Neural Networks,” *ICLR*, 2016  
[102] B. Recht, C. Re, S. Wright, and F. Niu ,“Hogwild: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent,” *NIPS*, 2011  
[103] D.J. Rezende, S. Mohamed, and D. Wierstra ,“Stochastic Backpropagation and Approximate Inference in Deep Generative Models,” *ICML*, 2014  
[104] M. Riedmiller ,“Neural Fitted Q Iteration---First Experiences with a Data Efficient Neural Reinforcement Learning Method,” *ECML*, 2005  
[105] S. Ross, G.J. Gordon, and D. Bagnell ,“A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning,” *AISTATS*, 2011  
[106] D.E. Rumelhart, G.E. Hinton, and R.J. Williams ,“Learning Representations by Back-Propagating Errors,” *Cognitive Modeling*, vol.5, iss.3, 1988  
[107] G.A. Rummery and M. Niranjan ,“{On-line Q-learning using Connectionist Systems,”  *University of Cambridge, Department of Engineering, 1994.*  
[108] A.A. Rusu, S.G. Colmenarejo, C. Gulcehre, G. Desjardins, J. Kirkpatrick, R. Pascanu, V. Mnih, K. Kavukcuoglu, and R. Hadsell ,“Policy Distillation,” *ICLR*, 2016  
[109] A.A. Rusu, N.C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell ,“Progressive Neural Networks,” *arXiv:1606.04671*, 2016  
[110] A.A. Rusu, M. Vecerik, T. Rothörl, N. Heess, R. Pascanu, and R. Hadsell ,“Sim-to-Real Robot Learning from Pixels with Progressive Nets,” *CoRL*, 2017  
[111] T. Salimans, J. Ho, X. Chen, and I. Sutskever ,“Evolution Strategies as a Scalable Alternative to Reinforcement Learning,” *arXiv:1703.03864*, 2017  
[112] T. Schaul, D. Horgan, K. Gregor, and D. Silver ,“Universal Value Function Approximators,” *ICML*, 2015  
[113] T. Schaul, J. Quan, I. Antonoglou, and D. Silver ,“Prioritized Experience Replay,” *ICLR*, 2016  
[114] J. Schmidhuber ,“A Possibility for Implementing Curiosity and Boredom in Model-Building Neural Controllers,” *SAB*, 1991  
[115] J. Schmidhuber and R. Huber ,“Learning to Generate Artificial Fovea Trajectories for Target Detection,” *IJNS*, vol.2, iss.01n02, pp.125-134, 1991  
[116] J. Schulman, N. Heess, T. Weber, and P. Abbeel ,“Gradient Estimation using Stochastic Computation Graphs,” *NIPS*, 2015  
[117] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz ,“Trust Region Policy Optimization,” *ICML*, 2015  
[118] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel ,“High-Dimensional Continuous Control using Generalized Advantage Estimation,” *ICLR*, 2016  
[119] J. Schulman, P. Abbeel, and X. Chen ,“Equivalence Between Policy Gradients and Soft Q-Learning,” *arXiv:1704.06440*, 2017  
[120] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov ,“Proximal Policy Optimization Algorithms,” *arXiv:1707.06347*, 2017  
[121] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, and N. Freitas ,“Taking the Human out of the Loop: A Review of Bayesian Optimization,” *Proc. of the IEEE*, vol.104, iss.1, pp.148-175, 2016  
[122] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller ,“Deterministic Policy Gradient Algorithms,” *ICML*, 2014  
[123] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, and et~al ,“Mastering the Game of Go with Deep Neural Networks and Tree Search,” *Nature*, vol.529, iss.7587, pp.484-489, 2016  
[124] S. Singh, D. Litman, M. Kearns, and M. Walker ,“Optimizing Dialogue Management with Reinforcement Learning: Experiments with the NJFun System,” *JAIR*, vol.16:, pp.105-133, 2002  
[125] I. Sorokin, A. Seleznev, M. Pavlov, A. Fedorov, and A. Ignateva ,“Deep Attention Recurrent Q-Network,” *NIPS Workshop on Deep Reinforcement Learning*, 2015  
[126] B.C. Stadie, P. Abbeel, and I. Sutskever ,“Third Person Imitation Learning,” *ICLR*, 2017  
[127] B.C. Stadie, S. Levine, and P. Abbeel ,“Incentivizing Exploration in Reinforcement Learning with Deep Predictive Models,” *NIPS Workshop on Deep Reinforcement Learning*, 2015  
[128] A.L. Strehl, L. Li, E. Wiewiora, J. Langford, and M.L. Littman ,“PAC Model-Free Reinforcement Learning,” *ICML*, 2006  
[129] S. Sukhbaatar, A. Szlam, and R. Fergus ,“Learning Multiagent Communication with Backpropagation,” *NIPS*, 2016  
[130] R.S. Sutton and A.G. Barto ,“{Reinforcement Learning: An Introduction,”  *MIT Press, 1998.*  
[131] R.S. Sutton, D. Precup, and S. Singh ,“Between MDPs and Semi-MDPs: A Framework for Temporal Abstraction in Reinforcement Learning,” *Artificial intelligence*, vol.112, iss.1--2, pp. 181-211, 1999  
[132] G. Synnaeve, N. Nardelli, A. Auvolat, S. Chintala, T. Lacroix, Z. Lin, F. Richoux, and N. Usunier ,“TorchCraft: A Library for Machine Learning Research on Real-Time Strategy Games,” *arXiv:1611.00625*, 2016  
[133] A. Tamar, Y. Wu, G. Thomas, S. Levine, and P. Abbeel ,“Value Iteration Networks,” *NIPS*, 2016  
[134] H. Tang, R. Houthooft, D. Foote, A. Stooke, X. Chen, Y. Duan, J. Schulman, F.D. Turck, and P. Abbeel ,“\#Exploration: A Study of Count-Based Exploration for Deep Reinforcement Learning,” *NIPS*, 2017  
[135] Y.W. Teh, V. Bapst, W.M. Czarnecki, J. Quan, J. Kirkpatrick, R. Hadsell, N. Heess, and R. Pascanu ,“Distral: Robust Multitask Reinforcement Learning,” *NIPS*, 2017  
[136] G. Tesauro ,“Temporal Difference Learning and TD-Gammon,” *Communications of the ACM*, vol.38, iss.3, pp.58-68, 1995  
[137] G. Tesauro, R. Das, H. Chan, J. Kephart, D. Levine, F. Rawson, and C. Lefurgy ,“Managing Power Consumption and Performance of Computing Systems using Reinforcement Learning,” *NIPS*, 2008  
[138] C. Tessler, S. Givony, T. Zahavy, D.J. Mankowitz, and S. Mannor ,“A Deep Hierarchical Approach to Lifelong Learning in Minecraft,” *AAAI*, 2017  
[139] E. Todorov, T. Erez, and Y. Tassa ,“MuJoCo: A Physics Engine for Model-Based Control,” *IROS*, 2012  
[140] J.N. Tsitsiklis and B.V. Roy ,“Analysis of Temporal-Difference Learning with Function Approximation,” *NIPS*, 1997  
[141] E. Tzeng, C. Devin, J. Hoffman, C. Finn, X. Peng, S. Levine, K. Saenko, and T. Darrell ,“Towards Adapting Deep Visuomotor Representations from Simulated to Real Environments,” *WAFR*, 2016  
[142] N. Usunier, G. Synnaeve, Z. Lin, and S. Chintala ,“Episodic Exploration for Deep Deterministic Policies: An Application to StarCraft Micromanagement Tasks,” *ICLR*, 2017  
[143] H. Hasselt ,“Double Q-Learning,” *NIPS*, 2010  
[144] H. Hasselt, A. Guez, and D. Silver ,“Deep Reinforcement Learning with Double Q-Learning,” *AAAI*, 2016  
[145] H. Vanseijen and R. Sutton ,“A Deeper Look at Planning as Learning from Replay,” *ICML*, 2015  
[146] A. Vezhnevets, V. Mnih, S. Osindero, A. Graves, O. Vinyals, J. Agapiou, and K. Kavukcuoglu ,“Strategic Attentive Writer for Learning Macro-Actions,” *NIPS*, 2016  
[147] A.S. Vezhnevets, S. Osindero, T. Schaul, N. Heess, M. Jaderberg, D. Silver, and K. Kavukcuoglu ,“FeUdal Networks for Hierarchical Reinforcement Learning,” *ICML*, 2017  
[148] O. Vinyals, T. Ewalds, S. Bartunov, P. Georgiev, A.S. Vezhnevets, M. Yeo, A. Makhzani, H. Küttler, J. Agapiou, J. Schrittwieser, and et~al ,“StarCraft II: A New Challenge for Reinforcement Learning,” *arXiv:1708.04782*, 2017  
[149] N. Wahlström, T.B. Schön, and M.P. Deisenroth ,“Learning Deep Dynamical Models from Image Pixels,” *IFAC SYSID*, vol.48, iss.28, 2015  
[150] N. Wahlström, T.B. Schön, and M.P. Deisenroth ,“From Pixels to Torques: Policy Learning with Deep Dynamical Models,” *ICML Workshop on Deep Learning*, 2015  
[151] J.X. Wang, Z.K. Nelson, D. Tirumala, H. Soyer, J.Z. Leibo, R. Munos, C. Blundell, D. Kumaran, and M. Botvinick ,“Learning to Reinforcement Learn,” *CogSci*, 2017  
[152] Z. Wang, N. Freitas, and M. Lanctot ,“Dueling Network Architectures for Deep Reinforcement Learning,” *ICLR*, 2016  
[153] Z. Wang, V. Bapst, N. Heess, V. Mnih, R. Munos, K. Kavukcuoglu, and N. Freitas ,“Sample Efficient Actor-Critic with Experience Replay,” *ICLR*, 2017  
[154] C.J.C.H. Watkins and P. Dayan ,“Q-Learning,” *Machine Learning*, vol.8, iss.3-4, pp.279-292, 1992  
[155] M. Watter, J. Springenberg, J. Boedecker, and M. Riedmiller ,“Embed to Control: A Locally Linear Latent Dynamics Model for Control from Raw Images,” *NIPS*, 2015  
[156] T. Weber, S. Racanière, D.P. Reichert, L. Buesing, A. Guez, D.J. Rezende, A.P. Badia, O. Vinyals, N. Heess, Y. Li, and et~al ,“Imagination-Augmented Agents for Deep Reinforcement Learning,” *NIPS*, 2017  
[157] P.J. Werbos ,“Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences,”  *Technical report, Harvard University, Applied Mathematics, 1974.*  
[158] D. Wierstra, A. Förster, J. Peters, and J. Schmidhuber ,“Recurrent Policy Gradients,” *Logic Journal of the IGPL*, vol.18, iss.5, pp.620-634, 2010  
[159] R.J. Williams ,“Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning,” *Machine Learning*, vol.8, iss.3-4, pp.229-256, 1992  
[160] M. Wulfmeier, P. Ondruska, and I. Posner ,“Maximum Entropy Deep Inverse Reinforcement Learning,” *NIPS Workshop on Deep Reinforcement Learning*, 2015  
[161] K. Xu, J. Ba, R. Kiros, K. Cho, A.C. Courville, R. Salakhutdinov, R.S. Zemel, and Y. Bengio ,“Show, Attend and Tell: Neural Image Caption Generation with Visual Attention,” *ICML*, 2015  
[162] Y. Zhu, R. Mottaghi, E. Kolve, J.J. Lim, A. Gupta, L.F. Fei, and A. Farhadi ,“Target-Driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning,” *ICRA*, 2017  
[163] B. Zoph and Q.V. Le ,“Neural Architecture Search with Reinforcement Learning,” *ICLR*, 2017