An Introduction to Deep Reinforcement Learning

Other titles in Foundations and Trends® in Machine Learning

Non-convex Optimization for Machine Learningy
Prateek Jain and Purushottam Ka
ISBN: 978-1-68083-368-3

Kernel Mean Embedding of Distributions: A Review and Beyond
Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur and Bernhard Scholkopf
ISBN: 978-1-68083-288-4

Tensor Networks for Dimensionality Reduction and Large-scale Optimization: Part 1 Low-Rank Tensor Decompositions
Andrzej Cichocki, Anh-Huy Phan, Qibin Zhao, Namgil Lee, Ivan Oseledets, Masashi Sugiyama and Danilo P. Mandic
ISBN: 978-1-68083-222-8

Tensor Networks for Dimensionality Reduction and Large-scale Optimization: Part 2 Applications and Future Perspectives
Andrzej Cichocki, Anh-Huy Phan, Qibin Zhao, Namgil Lee, Ivan Oseledets, Masashi Sugiyama and Danilo P. Mandic
ISBN: 978-1-68083-276-1

Patterns of Scalable Bayesian Inference
Elaine Angelino, Matthew James Johnson and Ryan P. Adams
ISBN: 978-1-68083-218-1

Generalized Low Rank Models
Madeleine Udell, Corinne Horn, Reza Zadeh and Stephen Boyd
ISBN: 978-1-68083-140-5

An Introduction to Deep Reinforcement Learning

Vincent François-Lavet
McGill University
vincent.francois-lavet@mcgill.ca

Peter Henderson
McGill University
peter.henderson@mail.mcgill.ca

Riashat Islam
McGill University
riashat.islam@mail.mcgill.ca

Marc G. Bellemare
Google Brain
bellemare@google.com

Joelle Pineau
Facebook, McGill University
jpineau@cs.mcgill.ca

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare and J. Pineau. *An Introduction to Deep Reinforcement Learning*. Foundations and Trends® in Machine Learning, vol. 11, no. 3-4, pp. 219–354, 2018.

ISBN: 978-1-68083-539-7

Foundations and Trends® in Machine Learning
Volume 11, Issue 3-4, 2018

Editorial Scope

Topics

Foundations and Trends® in Machine Learning publishes survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2018, Volume 11, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Contents

An Introduction to Deep Reinforcement Learning

Vincent François-Lavet[1], Peter Henderson[2], Riashat Islam[3], Marc G. Bellemare[4] and Joelle Pineau[5]

[1] *McGill University; vincent.francois-lavet@mcgill.ca*
[2] *McGill University; peter.henderson@mail.mcgill.ca*
[3] *McGill University; riashat.islam@mail.mcgill.ca*
[4] *Google Brain; bellemare@google.com*
[5] *Facebook, McGill University; jpineau@cs.mcgill.ca*

ABSTRACT

Deep reinforcement learning is the combination of reinforcement learning (RL) and deep learning. This field of research has been able to solve a wide range of complex decision-making tasks that were previously out of reach for a machine. Thus, deep RL opens up many new applications in domains such as healthcare, robotics, smart grids, finance, and many more. This manuscript provides an introduction to deep reinforcement learning models, algorithms and techniques. Particular focus is on the aspects related to generalization and how deep RL can be used for practical applications. We assume the reader is familiar with basic machine learning concepts.

Vincent François-Lavet, Peter Henderson, Riashat Islam, Marc G. Bellemare and Joelle Pineau (2018), "An Introduction to Deep Reinforcement Learning", Foundations and Trends® in Machine Learning: Vol. 11, No. 3-4, pp 219–354. DOI: 10.1561/2200000071.

1

Introduction

1.1 Motivation

A core topic in machine learning is that of sequential decision-making. This is the task of deciding, from experience, the sequence of actions to perform in an uncertain environment in order to achieve some goals. Sequential decision-making tasks cover a wide range of possible applications with the potential to impact many domains, such as robotics, healthcare, smart grids, finance, self-driving cars, and many more.

Inspired by behavioral psychology (see e.g., Sutton, 1984), reinforcement learning (RL) proposes a formal framework to this problem. The main idea is that an artificial agent may learn by interacting with its environment, similarly to a biological agent. Using the experience gathered, the artificial agent should be able to optimize some objectives given in the form of cumulative rewards. This approach applies in principle to any type of sequential decision-making problem relying on past experience. The environment may be stochastic, the agent may only observe partial information about the current state, the observations may be high-dimensional (e.g., frames and time series), the agent may freely gather experience in the environment or, on the contrary, the data may be may be constrained (e.g., not access to an accurate simulator or limited data).

Over the past few years, RL has become increasingly popular due to its success in addressing challenging sequential decision-making problems. Several of these achievements are due to the combination of RL with deep learning techniques (LeCun *et al.*, 2015; Schmidhuber, 2015; Goodfellow *et al.*, 2016). This combination, called deep RL, is most useful in problems with high dimensional state-space. Previous RL approaches had a difficult design issue in the choice of features (Munos and Moore, 2002; Bellemare *et al.*, 2013). However, deep RL has been successful in complicated tasks with lower prior knowledge thanks to its ability to learn different levels of abstractions from data. For instance, a deep RL agent can successfully learn from visual perceptual inputs made up of thousands of pixels (Mnih *et al.*, 2015). This opens up the possibility to mimic some human problem solving capabilities, even in high-dimensional space — which, only a few years ago, was difficult to conceive.

Several notable works using deep RL in games have stood out for attaining super-human level in playing Atari games from the pixels (Mnih *et al.*, 2015), mastering Go (Silver *et al.*, 2016b) or beating the world's top professionals at the game of Poker (Brown and Sandholm, 2017; Moravčik *et al.*, 2017). Deep RL also has potential for real-world applications such as robotics (Levine *et al.*, 2016; Gandhi *et al.*, 2017; Pinto *et al.*, 2017), self-driving cars (You *et al.*, 2017), finance (Deng *et al.*, 2017) and smart grids (François-Lavet, 2017), to name a few. Nonetheless, several challenges arise in applying deep RL algorithms. Among others, exploring the environment efficiently or being able to generalize a good behavior in a slightly different context are not straightforward. Thus, a large array of algorithms have been proposed for the deep RL framework, depending on a variety of settings of the sequential decision-making tasks.

1.2 Outline

The goal of this introduction to deep RL is to guide the reader towards effective use and understanding of core methods, as well as provide references for further reading. After reading this introduction, the reader should be able to understand the key different deep RL approaches and algorithms and should be able to apply them. The reader should also have enough background to investigate the scientific literature further and pursue research on deep RL.

In Chapter 2, we introduce the field of machine learning and the deep learning approach. The goal is to provide the general technical context and explain briefly where deep learning is situated in the broader field of machine learning. We assume the reader is familiar with basic notions of supervised and unsupervised learning; however, we briefly review the essentials.

In Chapter 3, we provide the general RL framework along with the case of a Markov Decision Process (MDP). In that context, we examine the different methodologies that can be used to train a deep RL agent. On the one hand, learning a value function (Chapter 4) and/or a direct representation of the policy (Chapter 5) belong to the so-called model-free approaches. On the other hand, planning algorithms that can make use of a learned model of the environment belong to the so-called model-based approaches (Chapter 6).

We dedicate Chapter 7 to the notion of generalization in RL. Within either a model-based or a model-free approach, we discuss the importance of different elements: (i) feature selection, (ii) function approximator selection, (iii) modifying the objective function and (iv) hierarchical learning. In Chapter 8, we present the main challenges of using RL in the online setting. In particular, we discuss the exploration-exploitation dilemma and the use of a replay memory.

In Chapter 9, we provide an overview of different existing benchmarks for evaluation of RL algorithms. Furthermore, we present a set of best practices to ensure consistency and reproducibility of the results obtained on the different benchmarks.

In Chapter 10, we discuss more general settings than MDPs: (i) the Partially Observable Markov Decision Process (POMDP), (ii) the distribution of MDPs (instead of a given MDP) along with the notion of transfer learning, (iii) learning without explicit reward function and (iv) multi-agent systems. We provide descriptions of how deep RL can be used in these settings.

In Chapter 11, we present broader perspectives on deep RL. This includes a discussion on applications of deep RL in various domains, along with the successes achieved and remaining challenges (e.g. robotics, self driving cars, smart grids, healthcare, etc.). This also includes a brief discussion on the relationship between deep RL and neuroscience.

Finally, we provide a conclusion in Chapter 12 with an outlook on the future development of deep RL techniques, their future applications, as well as the societal impact of deep RL and artificial intelligence.

Bahdanau, D., P. Brakel, K. Xu, A. Goyal, R. Lowe, J. Pineau, A. Courville, and Y. Bengio. 2016. "An actor-critic algorithm for sequence prediction". *arXiv preprint arXiv:1607.07086*.

Baird, L. 1995. "Residual algorithms: Reinforcement learning with function approximation". In: *ICML*. 30–37.

Baker, M. 2016. "1,500 scientists lift the lid on reproducibility". *Nature News*. 533(7604): 452.

Bartlett, P. L. and S. Mendelson. 2002. "Rademacher and Gaussian complexities: Risk bounds and structural results". *Journal of Machine Learning Research*. 3(Nov): 463–482.

Barto, A. G., R. S. Sutton, and C. W. Anderson. 1983. "Neuronlike adaptive elements that can solve difficult learning control problems". *IEEE transactions on systems, man, and cybernetics*. (5): 834–846.

Beattie, C., J. Z. Leibo, D. Teplyashin, T. Ward, M. Wainwright, H. Küttler, A. Lefrancq, S. Green, V. Valdés, A. Sadik, *et al.* 2016. "DeepMind Lab". *arXiv preprint arXiv:1612.03801*.

Bellemare, M. G., P. S. Castro, C. Gelada, K. Saurabh, and S. Moitra. 2018. "Dopamine". https://github.com/google/dopamine.

Bellemare, M. G., W. Dabney, and R. Munos. 2017. "A distributional perspective on reinforcement learning". *arXiv preprint arXiv:1707.06887*.

Bellemare, M. G., Y. Naddaf, J. Veness, and M. Bowling. 2013. "The Arcade Learning Environment: An evaluation platform for general agents." *Journal of Artificial Intelligence Research*. 47: 253–279.

Bellemare, M. G., S. Srinivasan, G. Ostrovski, T. Schaul, D. Saxton, and R. Munos. 2016. "Unifying Count-Based Exploration and Intrinsic Motivation". *arXiv preprint arXiv:1606.01868*.

Bellman, R. 1957a. "A Markovian decision process". *Journal of Mathematics and Mechanics*: 679–684.

Bellman, R. 1957b. "Dynamic Programming".

Bellman, R. E. and S. E. Dreyfus. 1962. "Applied dynamic programming".

Bello, I., H. Pham, Q. V. Le, M. Norouzi, and S. Bengio. 2016. "Neural Combinatorial Optimization with Reinforcement Learning". *arXiv preprint arXiv:1611.09940*.

Bengio, Y. 2017. “The Consciousness Prior”. *arXiv preprint arXiv:1709.08568*.

Bengio, Y., D.-H. Lee, J. Bornschein, T. Mesnard, and Z. Lin. 2015. “Towards biologically plausible deep learning”. *arXiv preprint arXiv:1502.04156*.

Bengio, Y., J. Louradour, R. Collobert, and J. Weston. 2009. “Curriculum learning”. In: *Proceedings of the 26th annual international conference on machine learning*. ACM. 41–48.

Bennett, C. C. and K. Hauser. 2013. “Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach”. *Artificial intelligence in medicine*. 57(1): 9–19.

Bertsekas, D. P., D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas. 1995. *Dynamic programming and optimal control*. Vol. 1. No. 2. Athena scientific Belmont, MA.

Bojarski, M., D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, *et al.* 2016. “End to end learning for self-driving cars”. *arXiv preprint arXiv:1604.07316*.

Bostrom, N. 2017. *Superintelligence*. Dunod.

Bouckaert, R. R. 2003. “Choosing between two learning algorithms based on calibrated tests”. In: *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. 51–58.

Bouckaert, R. R. and E. Frank. 2004. “Evaluating the replicability of significance tests for comparing learning algorithms”. In: *PAKDD*. Springer. 3–12.

Boularias, A., J. Kober, and J. Peters. 2011. “Relative Entropy Inverse Reinforcement Learning.” In: *AISTATS*. 182–189.

Boyan, J. A. and A. W. Moore. 1995. “Generalization in reinforcement learning: Safely approximating the value function”. In: *Advances in neural information processing systems*. 369–376.

Brafman, R. I. and M. Tennenholtz. 2003. “R-max-a general polynomial time algorithm for near-optimal reinforcement learning”. *The Journal of Machine Learning Research*. 3: 213–231.

Branavan, S., N. Kushman, T. Lei, and R. Barzilay. 2012. “Learning high-level planning from text”. In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics. 126–135.

Braziunas, D. 2003. “POMDP solution methods”. *University of Toronto, Tech. Rep.*

Brockman, G., V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. 2016. “OpenAI Gym”.

Brown, N. and T. Sandholm. 2017. “Libratus: The Superhuman AI for No-Limit Poker”. *International Joint Conference on Artificial Intelligence (IJCAI-17)*.

Browne, C. B., E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton. 2012. “A survey of monte carlo tree search methods”. *IEEE Transactions on Computational Intelligence and AI in games*. 4(1): 1–43.

Brügmann, B. 1993. “Monte carlo go”. *Tech. rep.* Citeseer.

Brundage, M., S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, P. Scharre, T. Zeitzoff, B. Filar, *et al.* 2018. “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation”. *arXiv preprint arXiv:1802.07228*.

Brys, T., A. Harutyunyan, P. Vrancx, M. E. Taylor, D. Kudenko, and A. Nowé. 2014. “Multi-objectivization of reinforcement learning problems by reward shaping”. In: *Neural Networks (IJCNN), 2014 International Joint Conference on*. IEEE. 2315–2322.

Bubeck, S., R. Munos, and G. Stoltz. 2011. “Pure exploration in finitely-armed and continuous-armed bandits”. *Theoretical Computer Science*. 412(19): 1832–1852.

Burda, Y., H. Edwards, A. Storkey, and O. Klimov. 2018. “Exploration by Random Network Distillation”. *arXiv preprint arXiv:1810.12894*.

Camerer, C., G. Loewenstein, and D. Prelec. 2005. “Neuroeconomics: How neuroscience can inform economics”. *Journal of economic Literature*. 43(1): 9–64.

Campbell, M., A. J. Hoane, and F.-h. Hsu. 2002. “Deep blue”. *Artificial intelligence*. 134(1-2): 57–83.

Casadevall, A. and F. C. Fang. 2010. "Reproducible science".

Castronovo, M., V. François-Lavet, R. Fonteneau, D. Ernst, and A. Couëtoux. 2017. "Approximate Bayes Optimal Policy Search using Neural Networks". In: *9th International Conference on Agents and Artificial Intelligence (ICAART 2017)*.

Chebotar, Y., A. Handa, V. Makoviychuk, M. Macklin, J. Issac, N. Ratliff, and D. Fox. 2018. "Closing the Sim-to-Real Loop: Adapting Simulation Randomization with Real World Experience". *arXiv preprint arXiv:1810.05687*.

Chen, T., I. Goodfellow, and J. Shlens. 2015. "Net2net: Accelerating learning via knowledge transfer". *arXiv preprint arXiv:1511.05641*.

Chen, X., C. Liu, and D. Song. 2017. "Learning Neural Programs To Parse Programs". *arXiv preprint arXiv:1706.01284*.

Chiappa, S., S. Racaniere, D. Wierstra, and S. Mohamed. 2017. "Recurrent Environment Simulators". *arXiv preprint arXiv:1704.02254*.

Christiano, P., J. Leike, T. B. Brown, M. Martic, S. Legg, and D. Amodei. 2017. "Deep reinforcement learning from human preferences". *arXiv preprint arXiv:1706.03741*.

Christopher, M. B. 2006. *Pattern recognition and machine learning*. Springer.

Cohen, J. D., S. M. McClure, and J. Y. Angela. 2007. "Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration". *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 362(1481): 933–942.

Cortes, C. and V. Vapnik. 1995. "Support-vector networks". *Machine learning*. 20(3): 273–297.

Coumans, E., Y. Bai, *et al.* 2016. "Bullet". http://pybullet.org/.

Da Silva, B., G. Konidaris, and A. Barto. 2012. "Learning parameterized skills". *arXiv preprint arXiv:1206.6398*.

Dabney, W., M. Rowland, M. G. Bellemare, and R. Munos. 2017. "Distributional Reinforcement Learning with Quantile Regression". *arXiv preprint arXiv:1710.10044*.

Dayan, P. and N. D. Daw. 2008. "Decision theory, reinforcement learning, and the brain". *Cognitive, Affective, & Behavioral Neuroscience*. 8(4): 429–453.

Dayan, P. and Y. Niv. 2008. “Reinforcement learning: the good, the bad and the ugly”. *Current opinion in neurobiology*. 18(2): 185–196.

Dearden, R., N. Friedman, and D. Andre. 1999. “Model based Bayesian exploration”. In: *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. 150–159.

Dearden, R., N. Friedman, and S. Russell. 1998. “Bayesian Q-learning”.

Deisenroth, M. and C. E. Rasmussen. 2011. “PILCO: A model-based and data-efficient approach to policy search”. In: *Proceedings of the 28th International Conference on machine learning (ICML-11)*. 465–472.

Demšar, J. 2006. “Statistical comparisons of classifiers over multiple data sets”. *Journal of Machine learning research*. 7(Jan): 1–30.

Deng, Y., F. Bao, Y. Kong, Z. Ren, and Q. Dai. 2017. “Deep direct reinforcement learning for financial signal representation and trading”. *IEEE transactions on neural networks and learning systems*. 28(3): 653–664.

Dhariwal, P., C. Hesse, M. Plappert, A. Radford, J. Schulman, S. Sidor, and Y. Wu. 2017. “OpenAI Baselines”.

Dietterich, T. G. 1998. “Approximate statistical tests for comparing supervised classification learning algorithms”. *Neural computation*. 10(7): 1895–1923.

Dietterich, T. G. 2009. “Machine learning and ecosystem informatics: challenges and opportunities”. In: *Asian Conference on Machine Learning*. Springer. 1–5.

Dinculescu, M. and D. Precup. 2010. “Approximate predictive representations of partially observable systems”. In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 895–902.

Dosovitskiy, A. and V. Koltun. 2016. “Learning to act by predicting the future”. *arXiv preprint arXiv:1611.01779*.

Duan, Y., M. Andrychowicz, B. Stadie, J. Ho, J. Schneider, I. Sutskever, P. Abbeel, and W. Zaremba. 2017. “One-Shot Imitation Learning”. *arXiv preprint arXiv:1703.07326*.

Duan, Y., X. Chen, R. Houthooft, J. Schulman, and P. Abbeel. 2016a. “Benchmarking deep reinforcement learning for continuous control”. In: *International Conference on Machine Learning*. 1329–1338.

Duan, Y., J. Schulman, X. Chen, P. L. Bartlett, I. Sutskever, and P. Abbeel. 2016b. "RL$\hat{2}$: Fast Reinforcement Learning via Slow Reinforcement Learning". *arXiv preprint arXiv:1611.02779*.

Duchesne, L., E. Karangelos, and L. Wehenkel. 2017. "Machine learning of real-time power systems reliability management response". *PowerTech Manchester 2017 Proceedings*.

Džeroski, S., L. De Raedt, and K. Driessens. 2001. "Relational reinforcement learning". *Machine learning*. 43(1-2): 7–52.

Erhan, D., Y. Bengio, A. Courville, and P. Vincent. 2009. "Visualizing higher-layer features of a deep network". *University of Montreal*. 1341(3): 1.

Ernst, D., P. Geurts, and L. Wehenkel. 2005. "Tree-based batch mode reinforcement learning". In: *Journal of Machine Learning Research*. 503–556.

Farquhar, G., T. Rocktäschel, M. Igl, and S. Whiteson. 2017. "TreeQN and ATreeC: Differentiable Tree Planning for Deep Reinforcement Learning". *arXiv preprint arXiv:1710.11417*.

Fazel-Zarandi, M., S.-W. Li, J. Cao, J. Casale, P. Henderson, D. Whitney, and A. Geramifard. 2017. "Learning Robust Dialog Policies in Noisy Environments". *arXiv preprint arXiv:1712.04034*.

Finn, C., P. Abbeel, and S. Levine. 2017. "Model-agnostic meta-learning for fast adaptation of deep networks". *arXiv preprint arXiv:1703.03400*.

Finn, C., I. Goodfellow, and S. Levine. 2016a. "Unsupervised learning for physical interaction through video prediction". In: *Advances In Neural Information Processing Systems*. 64–72.

Finn, C., S. Levine, and P. Abbeel. 2016b. "Guided cost learning: Deep inverse optimal control via policy optimization". In: *Proceedings of the 33rd International Conference on Machine Learning*. Vol. 48.

Florensa, C., Y. Duan, and P. Abbeel. 2017. "Stochastic neural networks for hierarchical reinforcement learning". *arXiv preprint arXiv:1704.03012*.

Florensa, C., D. Held, X. Geng, and P. Abbeel. 2018. "Automatic goal generation for reinforcement learning agents". In: *International Conference on Machine Learning*. 1514–1523.

Foerster, J., R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch. 2018. "Learning with opponent-learning awareness". In: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems. 122–130.

Foerster, J., G. Farquhar, T. Afouras, N. Nardelli, and S. Whiteson. 2017a. "Counterfactual Multi-Agent Policy Gradients". *arXiv preprint arXiv:1705.08926*.

Foerster, J., N. Nardelli, G. Farquhar, P. Torr, P. Kohli, S. Whiteson, *et al.* 2017b. "Stabilising experience replay for deep multi-agent reinforcement learning". *arXiv preprint arXiv:1702.08887*.

Fonteneau, R., S. A. Murphy, L. Wehenkel, and D. Ernst. 2013. "Batch mode reinforcement learning based on the synthesis of artificial trajectories". *Annals of operations research*. 208(1): 383–416.

Fonteneau, R., L. Wehenkel, and D. Ernst. 2008. "Variable selection for dynamic treatment regimes: a reinforcement learning approach".

Fortunato, M., M. G. Azar, B. Piot, J. Menick, I. Osband, A. Graves, V. Mnih, R. Munos, D. Hassabis, O. Pietquin, *et al.* 2017. "Noisy networks for exploration". *arXiv preprint arXiv:1706.10295*.

Fox, R., A. Pakman, and N. Tishby. 2015. "Taming the noise in reinforcement learning via soft updates". *arXiv preprint arXiv:1512.08562*.

François-Lavet, V. 2017. "Contributions to deep reinforcement learning and its applications in smartgrids". *PhD thesis*. University of Liege, Belgium.

François-Lavet, V. *et al.* 2016. "DeeR". https://deer.readthedocs.io/.

François-Lavet, V., Y. Bengio, D. Precup, and J. Pineau. 2018. "Combined Reinforcement Learning via Abstract Representations". *arXiv preprint arXiv:1809.04506*.

François-Lavet, V., D. Ernst, and F. Raphael. 2017. "On overfitting and asymptotic bias in batch reinforcement learning with partial observability". *arXiv preprint arXiv:1709.07796*.

François-Lavet, V., R. Fonteneau, and D. Ernst. 2015. "How to Discount Deep Reinforcement Learning: Towards New Dynamic Strategies". *arXiv preprint arXiv:1512.02011*.

François-Lavet, V., D. Taralla, D. Ernst, and R. Fonteneau. 2016. "Deep Reinforcement Learning Solutions for Energy Microgrids Management". In: *European Workshop on Reinforcement Learning.*

Fukushima, K. and S. Miyake. 1982. "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition". In: *Competition and cooperation in neural nets*. Springer. 267–285.

Gal, Y. and Z. Ghahramani. 2016. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning". In: *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*. 1050–1059.

Gandhi, D., L. Pinto, and A. Gupta. 2017. "Learning to Fly by Crashing". *arXiv preprint arXiv:1704.05588*.

Garnelo, M., K. Arulkumaran, and M. Shanahan. 2016. "Towards Deep Symbolic Reinforcement Learning". *arXiv preprint arXiv:1609.05518*.

Gauci, J., E. Conti, Y. Liang, K. Virochsiri, Y. He, Z. Kaden, V. Narayanan, and X. Ye. 2018. "Horizon: Facebook's Open Source Applied Reinforcement Learning Platform". *arXiv preprint arXiv:1811.00260*.

Gelly, S., Y. Wang, R. Munos, and O. Teytaud. 2006. "Modification of UCT with patterns in Monte-Carlo Go".

Geman, S., E. Bienenstock, and R. Doursat. 1992. "Neural networks and the bias/variance dilemma". *Neural computation*. 4(1): 1–58.

Geramifard, A., C. Dann, R. H. Klein, W. Dabney, and J. P. How. 2015. "RLPy: A Value-Function-Based Reinforcement Learning Framework for Education and Research". *Journal of Machine Learning Research*. 16: 1573–1578.

Geurts, P., D. Ernst, and L. Wehenkel. 2006. "Extremely randomized trees". *Machine learning*. 63(1): 3–42.

Ghavamzadeh, M., S. Mannor, J. Pineau, A. Tamar, *et al.* 2015. "Bayesian reinforcement learning: A survey". *Foundations and Trends® in Machine Learning*. 8(5-6): 359–483.

Giusti, A., J. Guzzi, D. C. Cireşan, F.-L. He, J. P. Rodriguez, F. Fontana, M. Faessler, C. Forster, J. Schmidhuber, G. Di Caro, *et al.* 2016. "A machine learning approach to visual perception of forest trails for mobile robots". *IEEE Robotics and Automation Letters*. 1(2): 661–667.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep learning*. MIT Press.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2672–2680.

Gordon, G. J. 1996. "Stable fitted reinforcement learning". In: *Advances in neural information processing systems*. 1052–1058.

Gordon, G. J. 1999. "Approximate solutions to Markov decision processes". *Robotics Institute*: 228.

Graves, A., G. Wayne, and I. Danihelka. 2014. "Neural turing machines". *arXiv preprint arXiv:1410.5401*.

Gregor, K., D. J. Rezende, and D. Wierstra. 2016. "Variational Intrinsic Control". *arXiv preprint arXiv:1611.07507*.

Gruslys, A., M. G. Azar, M. G. Bellemare, and R. Munos. 2017. "The Reactor: A Sample-Efficient Actor-Critic Architecture". *arXiv preprint arXiv:1704.04651*.

Gu, S., E. Holly, T. Lillicrap, and S. Levine. 2017a. "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates". In: *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE. 3389–3396.

Gu, S., T. Lillicrap, Z. Ghahramani, R. E. Turner, and S. Levine. 2017b. "Q-Prop: Sample-Efficient Policy Gradient with An Off-Policy Critic". In: *5th International Conference on Learning Representations (ICLR 2017)*.

Gu, S., T. Lillicrap, Z. Ghahramani, R. E. Turner, and S. Levine. 2016a. "Q-prop: Sample-efficient policy gradient with an off-policy critic". *arXiv preprint arXiv:1611.02247*.

Gu, S., T. Lillicrap, Z. Ghahramani, R. E. Turner, B. Schölkopf, and S. Levine. 2017c. "Interpolated Policy Gradient: Merging On-Policy and Off-Policy Gradient Estimation for Deep Reinforcement Learning". *arXiv preprint arXiv:1706.00387*.

Gu, S., T. Lillicrap, I. Sutskever, and S. Levine. 2016b. "Continuous Deep Q-Learning with Model-based Acceleration". *arXiv preprint arXiv:1603.00748*.

Guo, Z. D. and E. Brunskill. 2017. "Sample efficient feature selection for factored mdps". *arXiv preprint arXiv:1703.03454*.

Haarnoja, T., H. Tang, P. Abbeel, and S. Levine. 2017. "Reinforcement learning with deep energy-based policies". *arXiv preprint arXiv:1702.08165*.

Haber, N., D. Mrowca, L. Fei-Fei, and D. L. Yamins. 2018. "Learning to Play with Intrinsically-Motivated Self-Aware Agents". *arXiv preprint arXiv:1802.07442*.

Hadfield-Menell, D., S. J. Russell, P. Abbeel, and A. Dragan. 2016. "Cooperative inverse reinforcement learning". In: *Advances in neural information processing systems*. 3909–3917.

Hafner, R. and M. Riedmiller. 2011. "Reinforcement learning in feedback control". *Machine learning*. 84(1-2): 137–169.

Halsey, L. G., D. Curran-Everett, S. L. Vowler, and G. B. Drummond. 2015. "The fickle P value generates irreproducible results". *Nature methods*. 12(3): 179–185.

Harari, Y. N. 2014. *Sapiens: A brief history of humankind*.

Harutyunyan, A., M. G. Bellemare, T. Stepleton, and R. Munos. 2016. "Q (\lambda) with Off-Policy Corrections". In: *International Conference on Algorithmic Learning Theory*. Springer. 305–320.

Hassabis, D., D. Kumaran, C. Summerfield, and M. Botvinick. 2017. "Neuroscience-inspired artificial intelligence". *Neuron*. 95(2): 245–258.

Hasselt, H. V. 2010. "Double Q-learning". In: *Advances in Neural Information Processing Systems*. 2613–2621.

Hausknecht, M. and P. Stone. 2015. "Deep recurrent Q-learning for partially observable MDPs". *arXiv preprint arXiv:1507.06527*.

Hauskrecht, M., N. Meuleau, L. P. Kaelbling, T. Dean, and C. Boutilier. 1998. "Hierarchical solution of Markov decision processes using macro-actions". In: *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. 220–229.

He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep residual learning for image recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.

Heess, N., G. Wayne, D. Silver, T. Lillicrap, T. Erez, and Y. Tassa. 2015. "Learning continuous control policies by stochastic value gradients". In: *Advances in Neural Information Processing Systems*. 2944–2952.

Henderson, P., W.-D. Chang, F. Shkurti, J. Hansen, D. Meger, and G. Dudek. 2017a. "Benchmark Environments for Multitask Learning in Continuous Domains". *ICML Lifelong Learning: A Reinforcement Learning Approach Workshop*.

Henderson, P., R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger. 2017b. "Deep Reinforcement Learning that Matters". *arXiv preprint arXiv:1709.06560*.

Hessel, M., J. Modayil, H. van Hasselt, T. Schaul, G. Ostrovski, W. Dabney, D. Horgan, B. Piot, M. Azar, and D. Silver. 2017. "Rainbow: Combining Improvements in Deep Reinforcement Learning". *arXiv preprint arXiv:1710.02298*.

Hessel, M., H. Soyer, L. Espeholt, W. Czarnecki, S. Schmitt, and H. van Hasselt. 2018. "Multi-task Deep Reinforcement Learning with PopArt". *arXiv preprint arXiv:1809.04474*.

Higgins, I., A. Pal, A. A. Rusu, L. Matthey, C. P. Burgess, A. Pritzel, M. Botvinick, C. Blundell, and A. Lerchner. 2017. "Darla: Improving zero-shot transfer in reinforcement learning". *arXiv preprint arXiv:1707.08475*.

Ho, J. and S. Ermon. 2016. "Generative adversarial imitation learning". In: *Advances in Neural Information Processing Systems*. 4565–4573.

Hochreiter, S. and J. Schmidhuber. 1997. "Long short-term memory". *Neural computation*. 9(8): 1735–1780.

Hochreiter, S., A. S. Younger, and P. R. Conwell. 2001. "Learning to learn using gradient descent". In: *International Conference on Artificial Neural Networks*. Springer. 87–94.

Holroyd, C. B. and M. G. Coles. 2002. "The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity." *Psychological review*. 109(4): 679.

Houthooft, R., X. Chen, Y. Duan, J. Schulman, F. De Turck, and P. Abbeel. 2016. "Vime: Variational information maximizing exploration". In: *Advances in Neural Information Processing Systems*. 1109–1117.

Ioffe, S. and C. Szegedy. 2015. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". *arXiv preprint arXiv:1502.03167*.

Islam, R., P. Henderson, M. Gomrokchi, and D. Precup. 2017. "Reproducibility of Benchmarked Deep Reinforcement Learning Tasks for Continuous Control". *ICML Reproducibility in Machine Learning Workshop*.

Jaderberg, M., W. M. Czarnecki, I. Dunning, L. Marris, G. Lever, A. G. Castaneda, C. Beattie, N. C. Rabinowitz, A. S. Morcos, A. Ruderman, *et al.* 2018. "Human-level performance in first-person multiplayer games with population-based deep reinforcement learning". *arXiv preprint arXiv:1807.01281*.

Jaderberg, M., V. Mnih, W. M. Czarnecki, T. Schaul, J. Z. Leibo, D. Silver, and K. Kavukcuoglu. 2016. "Reinforcement learning with unsupervised auxiliary tasks". *arXiv preprint arXiv:1611.05397*.

Jakobi, N., P. Husbands, and I. Harvey. 1995. "Noise and the reality gap: The use of simulation in evolutionary robotics". In: *European Conference on Artificial Life*. Springer. 704–720.

James, G. M. 2003. "Variance and bias for general loss functions". *Machine Learning*. 51(2): 115–135.

Jaques, N., A. Lazaridou, E. Hughes, C. Gulcehre, P. A. Ortega, D. Strouse, J. Z. Leibo, and N. de Freitas. 2018. "Intrinsic Social Motivation via Causal Influence in Multi-Agent RL". *arXiv preprint arXiv:1810.08647*.

Jaquette, S. C. *et al.* 1973. "Markov decision processes with a new optimality criterion: Discrete time". *The Annals of Statistics*. 1(3): 496–505.

Jiang, N., A. Kulesza, and S. Singh. 2015a. “Abstraction selection in model-based reinforcement learning”. In: *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*. 179–188.

Jiang, N., A. Kulesza, S. Singh, and R. Lewis. 2015b. “The Dependence of Effective Planning Horizon on Model Accuracy”. In: *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems. 1181–1189.

Jiang, N. and L. Li. 2016. “Doubly robust off-policy value evaluation for reinforcement learning”. In: *Proceedings of The 33rd International Conference on Machine Learning*. 652–661.

Johnson, J., B. Hariharan, L. van der Maaten, J. Hoffman, L. Fei-Fei, C. L. Zitnick, and R. Girshick. 2017. “Inferring and Executing Programs for Visual Reasoning”. *arXiv preprint arXiv:1705.03633*.

Johnson, M., K. Hofmann, T. Hutton, and D. Bignell. 2016. “The Malmo Platform for Artificial Intelligence Experimentation.” In: *IJCAI*. 4246–4247.

Juliani, A., V.-P. Berges, E. Vckay, Y. Gao, H. Henry, M. Mattar, and D. Lange. 2018. “Unity: A General Platform for Intelligent Agents”. *arXiv preprint arXiv:1809.02627*.

Kaelbling, L. P., M. L. Littman, and A. R. Cassandra. 1998. “Planning and acting in partially observable stochastic domains”. *Artificial intelligence*. 101(1): 99–134.

Kahneman, D. 2011. *Thinking, fast and slow*. Macmillan.

Kakade, S. 2001. “A Natural Policy Gradient”. In: *Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, December 3-8, 2001, Vancouver, British Columbia, Canada]*. 1531–1538.

Kakade, S., M. Kearns, and J. Langford. 2003. “Exploration in metric state spaces”. In: *ICML*. Vol. 3. 306–312.

Kalakrishnan, M., P. Pastor, L. Righetti, and S. Schaal. 2013. “Learning objective functions for manipulation”. In: *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE. 1331–1336.

Kalashnikov, D., A. Irpan, P. Pastor, J. Ibarz, A. Herzog, E. Jang, D. Quillen, E. Holly, M. Kalakrishnan, V. Vanhoucke, and S. Levine. 2018. "Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation". *arXiv preprint arXiv:1806.10293*.

Kalchbrenner, N., A. v. d. Oord, K. Simonyan, I. Danihelka, O. Vinyals, A. Graves, and K. Kavukcuoglu. 2016. "Video pixel networks". *arXiv preprint arXiv:1610.00527*.

Kansky, K., T. Silver, D. A. Mély, M. Eldawy, M. Lázaro-Gredilla, X. Lou, N. Dorfman, S. Sidor, S. Phoenix, and D. George. 2017. "Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics". *arXiv preprint arXiv:1706.04317*.

Kaplan, R., C. Sauer, and A. Sosa. 2017. "Beating Atari with Natural Language Guided Reinforcement Learning". *arXiv preprint arXiv:1704.05539*.

Kearns, M. and S. Singh. 2002. "Near-optimal reinforcement learning in polynomial time". *Machine Learning*. 49(2-3): 209–232.

Kempka, M., M. Wydmuch, G. Runc, J. Toczek, and W. Jaśkowski. 2016. "Vizdoom: A doom-based ai research platform for visual reinforcement learning". In: *Computational Intelligence and Games (CIG), 2016 IEEE Conference on*. IEEE. 1–8.

Kirkpatrick, J., R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, *et al.* 2016. "Overcoming catastrophic forgetting in neural networks". *arXiv preprint arXiv:1612.00796*.

Klambauer, G., T. Unterthiner, A. Mayr, and S. Hochreiter. 2017. "Self-Normalizing Neural Networks". *arXiv preprint arXiv:1706.02515*.

Kolter, J. Z. and A. Y. Ng. 2009. "Near-Bayesian exploration in polynomial time". In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM. 513–520.

Konda, V. R. and J. N. Tsitsiklis. 2000. "Actor-critic algorithms". In: *Advances in neural information processing systems*. 1008–1014.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 1097–1105.

Kroon, M. and S. Whiteson. 2009. "Automatic feature selection for model-based reinforcement learning in factored MDPs". In: *Machine Learning and Applications, 2009. ICMLA'09. International Conference on*. IEEE. 324–330.

Kulkarni, T. D., K. Narasimhan, A. Saeedi, and J. Tenenbaum. 2016. "Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation". In: *Advances in Neural Information Processing Systems*. 3675–3683.

Lample, G. and D. S. Chaplot. 2017. "Playing FPS Games with Deep Reinforcement Learning." In: *AAAI*. 2140–2146.

LeCun, Y., Y. Bengio, *et al.* 1995. "Convolutional networks for images, speech, and time series". *The handbook of brain theory and neural networks*. 3361(10): 1995.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep learning". *Nature*. 521(7553): 436–444.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. "Gradient-based learning applied to document recognition". *Proceedings of the IEEE*. 86(11): 2278–2324.

Lee, D., H. Seo, and M. W. Jung. 2012. "Neural basis of reinforcement learning and decision making". *Annual review of neuroscience*. 35: 287–308.

Leffler, B. R., M. L. Littman, and T. Edmunds. 2007. "Efficient reinforcement learning with relocatable action models". In: *AAAI*. Vol. 7. 572–577.

Levine, S., C. Finn, T. Darrell, and P. Abbeel. 2016. "End-to-end training of deep visuomotor policies". *Journal of Machine Learning Research*. 17(39): 1–40.

Levine, S. and V. Koltun. 2013. "Guided policy search". In: *International Conference on Machine Learning*. 1–9.

Li, L., Y. Lv, and F.-Y. Wang. 2016. "Traffic signal timing via deep reinforcement learning". *IEEE/CAA Journal of Automatica Sinica*. 3(3): 247–254.

Li, L., W. Chu, J. Langford, and X. Wang. 2011. "Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms". In: *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM. 297–306.

Li, X., L. Li, J. Gao, X. He, J. Chen, L. Deng, and J. He. 2015. "Recurrent reinforcement learning: a hybrid approach". *arXiv preprint arXiv:1509.03044*.

Liaw, A., M. Wiener, *et al.* 2002. "Classification and regression by randomForest". *R news*. 2(3): 18–22.

Lillicrap, T. P., J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. 2015. "Continuous control with deep reinforcement learning". *arXiv preprint arXiv:1509.02971*.

Lin, L.-J. 1992. "Self-improving reactive agents based on reinforcement learning, planning and teaching". *Machine learning*. 8(3-4): 293–321.

Lipton, Z. C., J. Gao, L. Li, X. Li, F. Ahmed, and L. Deng. 2016. "Efficient exploration for dialogue policy learning with BBQ networks & replay buffer spiking". *arXiv preprint arXiv:1608.05081*.

Littman, M. L. 1994. "Markov games as a framework for multi-agent reinforcement learning". In: *Proceedings of the eleventh international conference on machine learning*. Vol. 157. 157–163.

Liu, Y., A. Gupta, P. Abbeel, and S. Levine. 2017. "Imitation from Observation: Learning to Imitate Behaviors from Raw Video via Context Translation". *arXiv preprint arXiv:1707.03374*.

Lowe, R., Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch. 2017. "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments". *arXiv preprint arXiv:1706.02275*.

MacGlashan, J., M. K. Ho, R. Loftin, B. Peng, D. Roberts, M. E. Taylor, and M. L. Littman. 2017. "Interactive Learning from Policy-Dependent Human Feedback". *arXiv preprint arXiv:1701.06049*.

Machado, M. C., M. G. Bellemare, and M. Bowling. 2017a. "A Laplacian Framework for Option Discovery in Reinforcement Learning". *arXiv preprint arXiv:1703.00956*.

Machado, M. C., M. G. Bellemare, E. Talvitie, J. Veness, M. Hausknecht, and M. Bowling. 2017b. "Revisiting the Arcade Learning Environment: Evaluation Protocols and Open Problems for General Agents". *arXiv preprint arXiv:1709.06009*.

Mandel, T., Y.-E. Liu, S. Levine, E. Brunskill, and Z. Popovic. 2014. "Offline policy evaluation across representations with applications to educational games". In: *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems. 1077–1084.

Mankowitz, D. J., T. A. Mann, and S. Mannor. 2016. "Adaptive Skills Adaptive Partitions (ASAP)". In: *Advances in Neural Information Processing Systems*. 1588–1596.

Mathieu, M., C. Couprie, and Y. LeCun. 2015. "Deep multi-scale video prediction beyond mean square error". *arXiv preprint arXiv:1511.05440*.

Matiisen, T., A. Oliver, T. Cohen, and J. Schulman. 2017. "Teacher-Student Curriculum Learning". *arXiv preprint arXiv:1707.00183*.

McCallum, A. K. 1996. "Reinforcement learning with selective perception and hidden state". *PhD thesis*. University of Rochester.

McGovern, A., R. S. Sutton, and A. H. Fagg. 1997. "Roles of macro-actions in accelerating reinforcement learning". In: *Grace Hopper celebration of women in computing*. Vol. 1317.

Miikkulainen, R., J. Liang, E. Meyerson, A. Rawal, D. Fink, O. Francon, B. Raju, A. Navruzyan, N. Duffy, and B. Hodjat. 2017. "Evolving Deep Neural Networks". *arXiv preprint arXiv:1703.00548*.

Mirowski, P., R. Pascanu, F. Viola, H. Soyer, A. Ballard, A. Banino, M. Denil, R. Goroshin, L. Sifre, K. Kavukcuoglu, *et al.* 2016. "Learning to navigate in complex environments". *arXiv preprint arXiv:1611.03673*.

Mnih, V., A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu. 2016. "Asynchronous methods for deep reinforcement learning". In: *International Conference on Machine Learning*.

Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.* 2015. "Human-level control through deep reinforcement learning". *Nature*. 518(7540): 529–533.

Mohamed, S. and D. J. Rezende. 2015. "Variational information maximisation for intrinsically motivated reinforcement learning". In: *Advances in neural information processing systems*. 2125–2133.

Montague, P. R. 2013. "Reinforcement Learning Models Then-and-Now: From Single Cells to Modern Neuroimaging". In: *20 Years of Computational Neuroscience*. Springer. 271–277.

Moore, A. W. 1990. "Efficient memory-based learning for robot control".

Morari, M. and J. H. Lee. 1999. "Model predictive control: past, present and future". *Computers & Chemical Engineering*. 23(4-5): 667–682.

Moravčik, M., M. Schmid, N. Burch, V. Lisy, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, and M. Bowling. 2017. "DeepStack: Expert-level artificial intelligence in heads-up no-limit poker". *Science*. 356(6337): 508–513.

Mordatch, I., K. Lowrey, G. Andrew, Z. Popovic, and E. V. Todorov. 2015. "Interactive control of diverse complex characters with neural networks". In: *Advances in Neural Information Processing Systems*. 3132–3140.

Morimura, T., M. Sugiyama, H. Kashima, H. Hachiya, and T. Tanaka. 2010. "Nonparametric return distribution approximation for reinforcement learning". In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 799–806.

Munos, R. and A. Moore. 2002. "Variable resolution discretization in optimal control". *Machine learning*. 49(2): 291–323.

Munos, R., T. Stepleton, A. Harutyunyan, and M. Bellemare. 2016. "Safe and efficient off-policy reinforcement learning". In: *Advances in Neural Information Processing Systems*. 1046–1054.

Murphy, K. P. 2012. "Machine Learning: A Probabilistic Perspective."

Nagabandi, A., G. Kahn, R. S. Fearing, and S. Levine. 2017. "Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning". *arXiv preprint arXiv:1708.02596*.

Nagabandi, A., G. Kahn, R. S. Fearing, and S. Levine. 2018. "Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning". In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 7559–7566.

Narvekar, S., J. Sinapov, M. Leonetti, and P. Stone. 2016. "Source task creation for curriculum learning". In: *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems. 566–574.

Neelakantan, A., Q. V. Le, and I. Sutskever. 2015. "Neural programmer: Inducing latent programs with gradient descent". *arXiv preprint arXiv:1511.04834*.

Neu, G. and C. Szepesvári. 2012. "Apprenticeship learning using inverse reinforcement learning and gradient methods". *arXiv preprint arXiv:1206.5264*.

Ng, A. Y., D. Harada, and S. Russell. 1999. "Policy invariance under reward transformations: Theory and application to reward shaping". In: *ICML*. Vol. 99. 278–287.

Ng, A. Y., S. J. Russell, *et al.* 2000. "Algorithms for inverse reinforcement learning." In: *Icml*. 663–670.

Nguyen, D. H. and B. Widrow. 1990. "Neural networks for self-learning control systems". *IEEE Control systems magazine*. 10(3): 18–23.

Niv, Y. 2009. "Reinforcement learning in the brain". *Journal of Mathematical Psychology*. 53(3): 139–154.

Niv, Y. and P. R. Montague. 2009. "Theoretical and empirical studies of learning". In: *Neuroeconomics*. Elsevier. 331–351.

Norris, J. R. 1998. *Markov chains*. No. 2. Cambridge university press.

O'Donoghue, B., R. Munos, K. Kavukcuoglu, and V. Mnih. 2016. "PGQ: Combining policy gradient and Q-learning". *arXiv preprint arXiv:1611.01626*.

Oh, J., V. Chockalingam, S. Singh, and H. Lee. 2016. "Control of Memory, Active Perception, and Action in Minecraft". *arXiv preprint arXiv:1605.09128*.

Oh, J., X. Guo, H. Lee, R. L. Lewis, and S. Singh. 2015. "Action-conditional video prediction using deep networks in atari games". In: *Advances in Neural Information Processing Systems*. 2863–2871.

Oh, J., S. Singh, and H. Lee. 2017. "Value Prediction Network". *arXiv preprint arXiv:1707.03497*.

Olah, C., A. Mordvintsev, and L. Schubert. 2017. "Feature Visualization". *Distill*. https://distill.pub/2017/feature-visualization.

Ortner, R., O.-A. Maillard, and D. Ryabko. 2014. "Selecting near-optimal approximate state representations in reinforcement learning". In: *International Conference on Algorithmic Learning Theory*. Springer. 140–154.

Osband, I., C. Blundell, A. Pritzel, and B. Van Roy. 2016. "Deep Exploration via Bootstrapped DQN". *arXiv preprint arXiv:1602.04621*.

Ostrovski, G., M. G. Bellemare, A. v. d. Oord, and R. Munos. 2017. "Count-based exploration with neural density models". *arXiv preprint arXiv:1703.01310*.

Paine, T. L., S. G. Colmenarejo, Z. Wang, S. Reed, Y. Aytar, T. Pfaff, M. W. Hoffman, G. Barth-Maron, S. Cabi, D. Budden, *et al.* 2018. "One-Shot High-Fidelity Imitation: Training Large-Scale Deep Nets with RL". *arXiv preprint arXiv:1810.05017*.

Parisotto, E., J. L. Ba, and R. Salakhutdinov. 2015. "Actor-mimic: Deep multitask and transfer reinforcement learning". *arXiv preprint arXiv:1511.06342*.

Pascanu, R., Y. Li, O. Vinyals, N. Heess, L. Buesing, S. Racanière, D. Reichert, T. Weber, D. Wierstra, and P. Battaglia. 2017. "Learning model-based planning from scratch". *arXiv preprint arXiv:1707.06170*.

Pathak, D., P. Agrawal, A. A. Efros, and T. Darrell. 2017. "Curiosity-driven exploration by self-supervised prediction". In: *International Conference on Machine Learning (ICML)*. Vol. 2017.

Pavlov, I. P. 1927. *Conditioned reflexes*. Oxford University Press.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.* 2011. "Scikit-learn: Machine learning in Python". *Journal of Machine Learning Research*. 12(Oct): 2825–2830.

Peng, J. and R. J. Williams. 1994. "Incremental multi-step Q-learning". In: *Machine Learning Proceedings 1994*. Elsevier. 226–232.

Peng, P., Q. Yuan, Y. Wen, Y. Yang, Z. Tang, H. Long, and J. Wang. 2017a. "Multiagent Bidirectionally-Coordinated Nets for Learning to Play StarCraft Combat Games". *arXiv preprint arXiv:1703.10069*.

Peng, X. B., G. Berseth, K. Yin, and M. van de Panne. 2017b. "DeepLoco: Dynamic Locomotion Skills Using Hierarchical Deep Reinforcement Learning". *ACM Transactions on Graphics (Proc. SIGGRAPH 2017)*. 36(4).

Perez-Liebana, D., S. Samothrakis, J. Togelius, T. Schaul, S. M. Lucas, A. Couëtoux, J. Lee, C.-U. Lim, and T. Thompson. 2016. "The 2014 general video game playing competition". *IEEE Transactions on Computational Intelligence and AI in Games*. 8(3): 229–243.

Petrik, M. and B. Scherrer. 2009. "Biasing approximate dynamic programming with a lower discount factor". In: *Advances in neural information processing systems*. 1265–1272.

Piketty, T. 2013. "Capital in the Twenty-First Century".

Pineau, J., G. Gordon, S. Thrun, *et al.* 2003. "Point-based value iteration: An anytime algorithm for POMDPs". In: *IJCAI*. Vol. 3. 1025–1032.

Pinto, L., M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel. 2017. "Asymmetric Actor Critic for Image-Based Robot Learning". *arXiv preprint arXiv:1710.06542*.

Plappert, M., R. Houthooft, P. Dhariwal, S. Sidor, R. Y. Chen, X. Chen, T. Asfour, P. Abbeel, and M. Andrychowicz. 2017. "Parameter Space Noise for Exploration". *arXiv preprint arXiv:1706.01905*.

Precup, D. 2000. "Eligibility traces for off-policy policy evaluation". *Computer Science Department Faculty Publication Series*: 80.

Ranzato, M., S. Chopra, M. Auli, and W. Zaremba. 2015. "Sequence level training with recurrent neural networks". *arXiv preprint arXiv:1511.06732*.

Rasmussen, C. E. 2004. "Gaussian processes in machine learning". In: *Advanced lectures on machine learning*. Springer. 63–71.

Ravindran, B. and A. G. Barto. 2004. "An algebraic approach to abstraction in reinforcement learning". *PhD thesis*. University of Massachusetts at Amherst.

Real, E., S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, Q. Le, and A. Kurakin. 2017. "Large-Scale Evolution of Image Classifiers". *arXiv preprint arXiv:1703.01041*.

Reed, S. and N. De Freitas. 2015. "Neural programmer-interpreters". *arXiv preprint arXiv:1511.06279*.

Rescorla, R. A., A. R. Wagner, *et al.* 1972. "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement". *Classical conditioning II: Current research and theory*. 2: 64–99.

Riedmiller, M. 2005. "Neural fitted Q iteration–first experiences with a data efficient neural reinforcement learning method". In: *Machine Learning: ECML 2005*. Springer. 317–328.

Riedmiller, M., R. Hafner, T. Lampe, M. Neunert, J. Degrave, T. Van de Wiele, V. Mnih, N. Heess, and J. T. Springenberg. 2018. "Learning by Playing - Solving Sparse Reward Tasks from Scratch". *arXiv preprint arXiv:1802.10567*.

Rowland, M., M. G. Bellemare, W. Dabney, R. Munos, and Y. W. Teh. 2018. "An Analysis of Categorical Distributional Reinforcement Learning". *arXiv preprint arXiv:1802.08163*.

Ruder, S. 2017. "An overview of multi-task learning in deep neural networks". *arXiv preprint arXiv:1706.05098*.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1988. "Learning representations by back-propagating errors". *Cognitive modeling*. 5(3): 1.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.* 2015. "Imagenet large scale visual recognition challenge". *International Journal of Computer Vision*. 115(3): 211–252.

Russek, E. M., I. Momennejad, M. M. Botvinick, S. J. Gershman, and N. D. Daw. 2017. "Predictive representations can link model-based reinforcement learning to model-free mechanisms". *bioRxiv*: 083857.

Rusu, A. A., S. G. Colmenarejo, C. Gulcehre, G. Desjardins, J. Kirkpatrick, R. Pascanu, V. Mnih, K. Kavukcuoglu, and R. Hadsell. 2015. "Policy distillation". *arXiv preprint arXiv:1511.06295*.

Rusu, A. A., M. Vecerik, T. Rothörl, N. Heess, R. Pascanu, and R. Hadsell. 2016. "Sim-to-real robot learning from pixels with progressive nets". *arXiv preprint arXiv:1610.04286*.

Sadeghi, F. and S. Levine. 2016. "CAD2RL: Real single-image flight without a single real image". *arXiv preprint arXiv:1611.04201*.

Salge, C., C. Glackin, and D. Polani. 2014. "Changing the environment based on empowerment as intrinsic motivation". *Entropy*. 16(5): 2789–2819.

Salimans, T., J. Ho, X. Chen, and I. Sutskever. 2017. "Evolution Strategies as a Scalable Alternative to Reinforcement Learning". *arXiv preprint arXiv:1703.03864*.

Samuel, A. L. 1959. "Some studies in machine learning using the game of checkers". *IBM Journal of research and development*. 3(3): 210–229.

Sandve, G. K., A. Nekrutenko, J. Taylor, and E. Hovig. 2013. "Ten simple rules for reproducible computational research". *PLoS computational biology*. 9(10): e1003285.

Santoro, A., D. Raposo, D. G. Barrett, M. Malinowski, R. Pascanu, P. Battaglia, and T. Lillicrap. 2017. "A simple neural network module for relational reasoning". *arXiv preprint arXiv:1706.01427*.

Savinov, N., A. Raichuk, R. Marinier, D. Vincent, M. Pollefeys, T. Lillicrap, and S. Gelly. 2018. "Episodic Curiosity through Reachability". *arXiv preprint arXiv:1810.02274*.

Schaarschmidt, M., A. Kuhnle, and K. Fricke. 2017. "TensorForce: A TensorFlow library for applied reinforcement learning".

Schaul, T., J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Sehnke, T. Rückstieß, and J. Schmidhuber. 2010. "PyBrain". *The Journal of Machine Learning Research*. 11: 743–746.

Schaul, T., D. Horgan, K. Gregor, and D. Silver. 2015a. "Universal value function approximators". In: *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*. 1312–1320.

Schaul, T., J. Quan, I. Antonoglou, and D. Silver. 2015b. "Prioritized Experience Replay". *arXiv preprint arXiv:1511.05952*.

Schmidhuber, J. 2010. "Formal theory of creativity, fun, and intrinsic motivation (1990–2010)". *IEEE Transactions on Autonomous Mental Development*. 2(3): 230–247.

Schmidhuber, J. 2015. "Deep learning in neural networks: An overview". *Neural Networks*. 61: 85–117.

Schraudolph, N. N., P. Dayan, and T. J. Sejnowski. 1994. "Temporal difference learning of position evaluation in the game of Go". In: *Advances in Neural Information Processing Systems*. 817–824.

Schulman, J., P. Abbeel, and X. Chen. 2017a. "Equivalence Between Policy Gradients and Soft Q-Learning". *arXiv preprint arXiv:1704.06440*.

Schulman, J., J. Ho, C. Lee, and P. Abbeel. 2016. "Learning from demonstrations through the use of non-rigid registration". In: *Robotics Research*. Springer. 339–354.

Schulman, J., S. Levine, P. Abbeel, M. I. Jordan, and P. Moritz. 2015. "Trust Region Policy Optimization". In: *ICML*. 1889–1897.

Schulman, J., F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. 2017b. "Proximal policy optimization algorithms". *arXiv preprint arXiv:1707.06347*.

Schultz, W., P. Dayan, and P. R. Montague. 1997. "A neural substrate of prediction and reward". *Science*. 275(5306): 1593–1599.

Shannon, C. 1950. "Programming a Computer for Playing Chess". *Philosophical Magazine*. 41(314).

Silver, D. L., Q. Yang, and L. Li. 2013. "Lifelong Machine Learning Systems: Beyond Learning Algorithms." In: *AAAI Spring Symposium: Lifelong Machine Learning*. Vol. 13. 05.

Silver, D., H. van Hasselt, M. Hessel, T. Schaul, A. Guez, T. Harley, G. Dulac-Arnold, D. Reichert, N. Rabinowitz, A. Barreto, *et al.* 2016a. "The predictron: End-to-end learning and planning". *arXiv preprint arXiv:1612.08810*.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.* 2016b. "Mastering the game of Go with deep neural networks and tree search". *Nature*. 529(7587): 484–489.

Silver, D., G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller. 2014. "Deterministic Policy Gradient Algorithms". In: *ICML*.

Singh, S. P., T. S. Jaakkola, and M. I. Jordan. 1994. "Learning Without State-Estimation in Partially Observable Markovian Decision Processes." In: *ICML*. 284–292.

Singh, S. P. and R. S. Sutton. 1996. "Reinforcement learning with replacing eligibility traces". *Machine learning*. 22(1-3): 123–158.

Singh, S., T. Jaakkola, M. L. Littman, and C. Szepesvári. 2000. "Convergence results for single-step on-policy reinforcement-learning algorithms". *Machine learning*. 38(3): 287–308.

Sondik, E. J. 1978. "The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs". *Operations research*. 26(2): 282–304.

Srivastava, N., G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. "Dropout: a simple way to prevent neural networks from overfitting." *Journal of Machine Learning Research*. 15(1): 1929–1958.

Stadie, B. C., S. Levine, and P. Abbeel. 2015. "Incentivizing Exploration In Reinforcement Learning With Deep Predictive Models". *arXiv preprint arXiv:1507.00814*.

Stone, P. and M. Veloso. 2000. "Layered learning". *Machine Learning: ECML 2000*: 369–381.

Story, G., I. Vlaev, B. Seymour, A. Darzi, and R. Dolan. 2014. "Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective". *Frontiers in behavioral neuroscience*. 8: 76.

Sukhbaatar, S., A. Szlam, and R. Fergus. 2016. "Learning multiagent communication with backpropagation". In: *Advances in Neural Information Processing Systems*. 2244–2252.

Sun, Y., F. Gomez, and J. Schmidhuber. 2011. "Planning to be surprised: Optimal bayesian exploration in dynamic environments". In: *Artificial General Intelligence*. Springer. 41–51.

Sunehag, P., G. Lever, A. Gruslys, W. M. Czarnecki, V. Zambaldi, M. Jaderberg, M. Lanctot, N. Sonnerat, J. Z. Leibo, K. Tuyls, *et al.* 2017. "Value-Decomposition Networks For Cooperative Multi-Agent Learning". *arXiv preprint arXiv:1706.05296*.

Sutton, R. S. 1988. "Learning to predict by the methods of temporal differences". *Machine learning*. 3(1): 9–44.

Sutton, R. S. 1996. "Generalization in reinforcement learning: Successful examples using sparse coarse coding". *Advances in neural information processing systems*: 1038–1044.

Sutton, R. S. and A. G. Barto. 1998. *Reinforcement learning: An introduction*. Vol. 1. No. 1. MIT press Cambridge.

Sutton, R. S. and A. G. Barto. 2017. *Reinforcement Learning: An Introduction (2nd Edition, in progress)*. MIT Press.

Sutton, R. S., D. A. McAllester, S. P. Singh, and Y. Mansour. 2000. "Policy gradient methods for reinforcement learning with function approximation". In: *Advances in neural information processing systems*. 1057–1063.

Sutton, R. S., D. Precup, and S. Singh. 1999. "Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning". *Artificial intelligence*. 112(1-2): 181–211.

Sutton, R. S. 1984. "Temporal credit assignment in reinforcement learning".

Synnaeve, G., N. Nardelli, A. Auvolat, S. Chintala, T. Lacroix, Z. Lin, F. Richoux, and N. Usunier. 2016. "TorchCraft: a Library for Machine Learning Research on Real-Time Strategy Games". *arXiv preprint arXiv:1611.00625*.

Szegedy, C., S. Ioffe, V. Vanhoucke, and A. Alemi. 2016. "Inception-v4, inception-resnet and the impact of residual connections on learning". *arXiv preprint arXiv:1602.07261*.

Szegedy, C., S. Ioffe, V. Vanhoucke, and A. A. Alemi. 2017. "Inception-v4, inception-resnet and the impact of residual connections on learning." In: *AAAI*. Vol. 4. 12.

Tamar, A., S. Levine, P. Abbeel, Y. WU, and G. Thomas. 2016. "Value iteration networks". In: *Advances in Neural Information Processing Systems*. 2146–2154.

Tan, J., T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke. 2018. "Sim-to-Real: Learning Agile Locomotion For Quadruped Robots". *arXiv preprint arXiv:1804.10332*.

Tanner, B. and A. White. 2009. "RL-Glue: Language-independent software for reinforcement-learning experiments". *The Journal of Machine Learning Research*. 10: 2133–2136.

Teh, Y. W., V. Bapst, W. M. Czarnecki, J. Quan, J. Kirkpatrick, R. Hadsell, N. Heess, and R. Pascanu. 2017. "Distral: Robust Multitask Reinforcement Learning". *arXiv preprint arXiv:1707.04175*.

Tesauro, G. 1995. "Temporal difference learning and TD-Gammon". *Communications of the ACM*. 38(3): 58–68.

Tessler, C., S. Givony, T. Zahavy, D. J. Mankowitz, and S. Mannor. 2017. "A Deep Hierarchical Approach to Lifelong Learning in Minecraft." In: *AAAI*. 1553–1561.

Thomas, P. 2014. “Bias in natural actor-critic algorithms”. In: *International Conference on Machine Learning*. 441–448.

Thomas, P. S. and E. Brunskill. 2016. “Data-efficient off-policy policy evaluation for reinforcement learning”. In: *International Conference on Machine Learning*.

Thrun, S. B. 1992. “Efficient exploration in reinforcement learning”.

Tian, Y., Q. Gong, W. Shang, Y. Wu, and C. L. Zitnick. 2017. “ELF: An Extensive, Lightweight and Flexible Research Platform for Real-time Strategy Games”. *Advances in Neural Information Processing Systems (NIPS)*.

Tieleman, H. 2012. “Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude”. *COURSERA: Neural Networks for Machine Learning*.

Tobin, J., R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel. 2017. “Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World”. *arXiv preprint arXiv:1703.06907*.

Todorov, E., T. Erez, and Y. Tassa. 2012. “MuJoCo: A physics engine for model-based control”. In: *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE. 5026–5033.

Tsitsiklis, J. N. and B. Van Roy. 1997. “An analysis of temporal-difference learning with function approximation”. *Automatic Control, IEEE Transactions on*. 42(5): 674–690.

Turing, A. M. 1953. “Digital computers applied to games”. *Faster than thought*.

Tzeng, E., C. Devin, J. Hoffman, C. Finn, P. Abbeel, S. Levine, K. Saenko, and T. Darrell. 2015. “Adapting deep visuomotor representations with weak pairwise constraints”. *arXiv preprint arXiv:1511.07111*.

Ueno, S., M. Osawa, M. Imai, T. Kato, and H. Yamakawa. 2017. ““Re: ROS”: Prototyping of Reinforcement Learning Environment for Asynchronous Cognitive Architecture”. In: *First International Early Research Career Enhancement School on Biologically Inspired Cognitive Architectures*. Springer. 198–203.

Van Hasselt, H., A. Guez, and D. Silver. 2016. “Deep Reinforcement Learning with Double Q-Learning.” In: *AAAI*. 2094–2100.

Vapnik, V. N. 1998. "Statistical learning theory. Adaptive and learning systems for signal processing, communications, and control".

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. 2017. "Attention Is All You Need". *arXiv preprint arXiv:1706.03762*.

Vezhnevets, A., V. Mnih, S. Osindero, A. Graves, O. Vinyals, J. Agapiou, *et al.* 2016. "Strategic attentive writer for learning macro-actions". In: *Advances in Neural Information Processing Systems*. 3486–3494.

Vinyals, O., T. Ewalds, S. Bartunov, P. Georgiev, A. S. Vezhnevets, M. Yeo, A. Makhzani, H. Küttler, J. Agapiou, J. Schrittwieser, *et al.* 2017. "StarCraft II: A New Challenge for Reinforcement Learning". *arXiv preprint arXiv:1708.04782*.

Wahlström, N., T. B. Schön, and M. P. Deisenroth. 2015. "From pixels to torques: Policy learning with deep dynamical models". *arXiv preprint arXiv:1502.02251*.

Walsh, T. 2017. *It's Alive!: Artificial Intelligence from the Logic Piano to Killer Robots*. La Trobe University Press.

Wang, J. X., Z. Kurth-Nelson, D. Tirumala, H. Soyer, J. Z. Leibo, R. Munos, C. Blundell, D. Kumaran, and M. Botvinick. 2016a. "Learning to reinforcement learn". *arXiv preprint arXiv:1611.05763*.

Wang, Z., V. Bapst, N. Heess, V. Mnih, R. Munos, K. Kavukcuoglu, and N. de Freitas. 2016b. "Sample efficient actor-critic with experience replay". *arXiv preprint arXiv:1611.01224*.

Wang, Z., N. de Freitas, and M. Lanctot. 2015. "Dueling network architectures for deep reinforcement learning". *arXiv preprint arXiv:1511.06581*.

Warnell, G., N. Waytowich, V. Lawhern, and P. Stone. 2017. "Deep TAMER: Interactive Agent Shaping in High-Dimensional State Spaces". *arXiv preprint arXiv:1709.10163*.

Watkins, C. J. and P. Dayan. 1992. "Q-learning". *Machine learning*. 8(3-4): 279–292.

Watkins, C. J. C. H. 1989. "Learning from delayed rewards". *PhD thesis*. King's College, Cambridge.

Watter, M., J. Springenberg, J. Boedecker, and M. Riedmiller. 2015. "Embed to control: A locally linear latent dynamics model for control from raw images". In: *Advances in neural information processing systems*. 2746–2754.

Weber, T., S. Racanière, D. P. Reichert, L. Buesing, A. Guez, D. J. Rezende, A. P. Badia, O. Vinyals, N. Heess, Y. Li, *et al.* 2017. "Imagination-Augmented Agents for Deep Reinforcement Learning". *arXiv preprint arXiv:1707.06203*.

Wender, S. and I. Watson. 2012. "Applying reinforcement learning to small scale combat in the real-time strategy game StarCraft: Broodwar". In: *Computational Intelligence and Games (CIG), 2012 IEEE Conference on*. IEEE. 402–408.

Whiteson, S., B. Tanner, M. E. Taylor, and P. Stone. 2011. "Protecting against evaluation overfitting in empirical reinforcement learning". In: *Adaptive Dynamic Programming And Reinforcement Learning (ADPRL), 2011 IEEE Symposium on*. IEEE. 120–127.

Williams, R. J. 1992. "Simple statistical gradient-following algorithms for connectionist reinforcement learning". *Machine learning*. 8(3-4): 229–256.

Wu, Y. and Y. Tian. 2016. "Training agent for first-person shooter game with actor-critic curriculum learning".

Xu, K., J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio. 2015. "Show, attend and tell: Neural image caption generation with visual attention". In: *International Conference on Machine Learning*. 2048–2057.

You, Y., X. Pan, Z. Wang, and C. Lu. 2017. "Virtual to Real Reinforcement Learning for Autonomous Driving". *arXiv preprint arXiv:1704.03952*.

Zamora, I., N. G. Lopez, V. M. Vilches, and A. H. Cordero. 2016. "Extending the OpenAI Gym for robotics: a toolkit for reinforcement learning using ROS and Gazebo". *arXiv preprint arXiv:1608.05742*.

Zhang, A., N. Ballas, and J. Pineau. 2018a. "A Dissection of Overfitting and Generalization in Continuous Reinforcement Learning". *arXiv preprint arXiv:1806.07937*.

Zhang, A., H. Satija, and J. Pineau. 2018b. "Decoupling Dynamics and Reward for Transfer Learning". *arXiv preprint arXiv:1804.10689*.

Zhang, C., O. Vinyals, R. Munos, and S. Bengio. 2018c. "A Study on Overfitting in Deep Reinforcement Learning". *arXiv preprint arXiv:1804.06893*.

Zhang, C., S. Bengio, M. Hardt, B. Recht, and O. Vinyals. 2016. "Understanding deep learning requires rethinking generalization". *arXiv preprint arXiv:1611.03530*.

Zhu, Y., R. Mottaghi, E. Kolve, J. J. Lim, A. Gupta, L. Fei-Fei, and A. Farhadi. 2016. "Target-driven visual navigation in indoor scenes using deep reinforcement learning". *arXiv preprint arXiv:1609.05143*.

Ziebart, B. D. 2010. *Modeling purposeful adaptive behavior with the principle of maximum causal entropy*. Carnegie Mellon University.

Zoph, B. and Q. V. Le. 2016. "Neural architecture search with reinforcement learning". *arXiv preprint arXiv:1611.01578*.