



## UvA-DARE (Digital Academic Repository)

### Combining strategies efficiently: high-quality decisions from conflicting advice

Koolen, W.M.

**Publication date**  
2011

[Link to publication](#)

#### **Citation for published version (APA):**

Koolen, W. M. (2011). *Combining strategies efficiently: high-quality decisions from conflicting advice*. [Thesis, fully internal, Universiteit van Amsterdam].

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

---

## Bibliography

- [1] N. Abe and P. M. Long. Associative reinforcement learning using linear probabilistic concepts. In *Proc. 16th International Conf. on Machine Learning*, pages 3–11. Morgan Kaufmann, San Francisco, CA, 1999.
- [2] J. Abernethy and M. K. Warmuth. Repeated games against budgeted adversaries. Unpublished manuscript.
- [3] J. Abernethy, M. K. Warmuth, and J. Yellin. When random play is optimal against an adversary. Journal version of [6], in progress.
- [4] J. Abernethy, J. Langford, and M. K. Warmuth. Continuous experts and the binning algorithm. In G. Lugosi and H. Simon, editors, *Learning Theory*, volume 4005 of *Lecture Notes in Computer Science*, pages 544–558. Springer Berlin / Heidelberg, 2006.
- [5] J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *In Proceedings of the 21st Annual Conference on Learning Theory (COLT, 2008)*.
- [6] J. Abernethy, M. K. Warmuth, and J. Yellin. Optimal strategies for random walks. In *Proceedings of The 21st Annual Conference on Learning Theory*, pages 437–446, July 2008.
- [7] R. Ananthanarayanan, S. K. Esser, H. D. Simon, and D. S. Modha. The cat is out of the bag: cortical simulations with  $10^9$  neurons,

- $10^{13}$  synapses. In *SC '09: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, pages 1–12, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-744-8.
- [8] M. Anthony and N. Biggs. *Computational Learning Theory*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, UK, 1992.
- [9] M. Asada, S. Noda, S. Tawaratsumida, and K. Hosoda. Purposive behavior acquisition for a real robot by vision-based reinforcement learning. *Machine Learning*, 23:279–303, 1996.
- [10] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2/3):235–256, 2002.
- [11] L. Baird. Residual algorithms: reinforcement learning with function approximation. In *Proc. 12th International Conference on Machine Learning*, pages 30–37. Morgan Kaufmann, 1995.
- [12] P. L. Bartlett and J. Baxter. Estimation and approximation bounds for gradient-based reinforcement learning. In *Proc. 13th Annu. Conference on Comput. Learning Theory*, pages 133–141. Morgan Kaufmann, San Francisco, 2000.
- [13] P. L. Bartlett and J. Baxter. Estimation and approximation bounds for gradient-based reinforcement learning. *J. Comput. Syst. Sci.*, 64(1):133–150, 2002. Special Issue for COLT 2000.
- [14] J. Baxter and P. L. Bartlett. Reinforcement learning in POMDP's via direct gradient ascent. In *Proc. 17th International Conf. on Machine Learning*, pages 41–48. Morgan Kaufmann, San Francisco, CA, 2000.
- [15] J. O. Berger. Could Fisher, Jeffreys and Neyman have agreed on testing? *Statistical Science*, 18(1):1–32, 2003.
- [16] K. Binmore. *Fun and games: a text on game theory*. D.C. Heath, 1991.
- [17] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.

- [18] O. Bousquet. A note on parameter tuning for on-line shifting algorithms. Technical report, Max Planck Institute for Biological Cybernetics, 2003.
- [19] O. Bousquet and M. K. Warmuth. Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research*, 3: 363–396, 2002.
- [20] M. Bowling. Convergence problems of general-sum multiagent reinforcement learning. In *Proc. 17th International Conf. on Machine Learning*, pages 89–94. Morgan Kaufmann, San Francisco, CA, 2000.
- [21] R. Bracewell. “Convolution” and “Two-dimensional convolution”. In *The Fourier Transform and Its Applications*, pages 25–50 and 243–244. McGraw-Hill, 1965.
- [22] R. I. Brafman and M. Tennenholtz. R-MAX - A general polynomial time algorithm for near-optimal reinforcement learning. *Journal of Machine Learning Research*, 3:213–231, 2002.
- [23] A. Broder. Generating random spanning trees. In *SFCS '89: Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, pages 442–447, Washington, DC, USA, 1989. IEEE Computer Society. ISBN 0-8186-1982-1.
- [24] N. Cesa-Bianchi and P. Fischer. Finite-time regret bounds for the multiarmed bandit problem. In *Proc. 15th International Conf. on Machine Learning*, pages 100–108. Morgan Kaufmann, San Francisco, CA, 1998.
- [25] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.
- [26] N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. In *Proceedings of the 22nd Annual Conference on Learning Theory*, 2009.
- [27] N. Cesa-Bianchi, Y. Freund, D. P. Helmbold, and M. K. Warmuth. On-line prediction and conversion strategies. *Machine Learning*, 25:71–110, 1996. An extended abstract appeared in *EuroColt '93*.

- [28] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, May 1997.
- [29] G.-H. Chen, M.-Y. Kao, Y.-D. Lyuu, and H.-K. Wong. Optimal buy-and-hold strategies for financial markets with bounded daily returns. In *Proc. of the 31st annual ACM symposium on Theory of computing*, pages 119–128. ACM, 1999. ISBN 1-58113-067-8.
- [30] P. Cichosz and J. J. Mulawka. Fast and efficient reinforcement learning with truncated temporal differences. In *Proc. 12th International Conference on Machine Learning*, pages 99–107. Morgan Kaufmann, 1995.
- [31] J. G. Cleary, Ian, and I. H. Witten. Data compression using adaptive coding and partial string matching. *IEEE Transactions on Communications*, 32:396–402, 1984.
- [32] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301, 1965. ISSN 00255718.
- [33] T. M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.
- [34] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.
- [35] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, Cambridge, UK, 2000.
- [36] R. H. Crites and A. G. Barto. Elevator group control using multiple reinforcement learning agents. *Machine Learning*, 33(2/3): 235–262, 1998.
- [37] F. A. Dahl. A reinforcement learning algorithm applied to simplified two-player texas hold'em poker. In *Machine Learning: ECML 2001, 12th European Conference on Machine Learning, Freiburg, Germany, September 5-7, 2001, Proceedings*, volume 2167 of *Lecture Notes in Artificial Intelligence*, pages 85–96. Springer, 2001.

- [38] E. Dannoura and K. Sakurai. An improvement on El-Yaniv-Fiat-Karp-Turpin's money-making bi-directional trading strategy. *IPL*, 66(1):27–33, 1998.
- [39] A. P. Dawid. Statistical theory: The prequential approach. *Journal of the Royal Statistical Society, Series A*, 147, Part 2:278–292, 1984.
- [40] A. P. Dawid, S. de Rooij, G. Shafer, A. Shen, N. Vereshchagin, and V. Vovk. Insuring against loss of evidence in game-theoretic probability. *Statistics & Probability Letters*, 81(1):157 – 162, 2011. ISSN 0167-7152.
- [41] S. de Rooij and T. van Erven. Learning the switching rate by discretising Bernoulli sources online. In *JMLR Workshop and Conference Proceedings*, volume 5: AISTATS, 2009.
- [42] G. DeJong. Hidden strengths and limitations: An empirical investigation of reinforcement learning. In *Proc. 17th International Conf. on Machine Learning*, pages 215–222. Morgan Kaufmann, San Francisco, CA, 2000.
- [43] P. DeMarzo, I. Kremer, and Y. Mansour. Online trading algorithms and robust option pricing. In *Proc. of the 38 annual ACM symposium on Theory of computing*, pages 477–486. ACM, 2006. ISBN 1-59593-134-1.
- [44] T. G. Dietterich. The MAXQ method for hierarchical reinforcement learning. In *Proc. 15th International Conf. on Machine Learning*, pages 118–126. Morgan Kaufmann, San Francisco, CA, 1998.
- [45] T. G. Dietterich. Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*, 13:227–303, 2000.
- [46] T. G. Dietterich and N. S. Flann. Explanation-based learning and reinforcement learning: a unified view. *Machine Learning*, 28:169–210, 1997. Earlier version in 12th International Conf on ML, 1995.
- [47] T. G. Dietterich and X. Wang. Support vectors for reinforcement learning. In *Machine Learning: ECML 2001, 12th European Conference on Machine Learning, Freiburg, Germany, September 5-7, 2001*,

- Proceedings*, volume 2167 of *Lecture Notes in Artificial Intelligence*, page 600. Springer, 2001.
- [48] C. Domingo. Faster near-optimal reinforcement learning: Adding adaptiveness to the  $E^3$  algorithm. In *Algorithmic Learning Theory, 10th International Conference, ALT '99, Tokyo, Japan, December 1999, Proceedings*, volume 1720 of *Lecture Notes in Artificial Intelligence*, pages 241–251. Springer, 1999.
- [49] K. Driessens, J. Ramon, and H. Blockeel. Speeding up relational reinforcement learning through the use of an incremental first order decision tree learner. In *Machine Learning: ECML 2001, 12th European Conference on Machine Learning, Freiburg, Germany, September 5-7, 2001, Proceedings*, volume 2167 of *Lecture Notes in Artificial Intelligence*, pages 97–108. Springer, 2001.
- [50] M. O. Duff. Q-learning for bandit problems. In *Proc. 12th International Conference on Machine Learning*, pages 209–217. Morgan Kaufmann, 1995.
- [51] S. Džeroski, L. De Raedt, and H. Blockeel. Relational reinforcement learning. In *Proc. 15th International Conf. on Machine Learning*, pages 136–143. Morgan Kaufmann, San Francisco, CA, 1998.
- [52] S. Dzeroski, L. De Raedt, and K. Driessens. Relational reinforcement learning. *Machine Learning*, 43(1/2):7–52, 2001.
- [53] R. El-Yaniv, A. Fiat, R. M. Karp, and G. Turpin. Optimal search and one-way trading online algorithms. *Algorithmica*, 30(1):101–139, 2001.
- [54] D. Ernst, P. Geurts, and L. Wehenkel. Tree-based batch mode reinforcement learning. *Journal of Machine Learning Research*, 6: 503–556, 2005.
- [55] E. Even-Dar, S. Mannor, and Y. Mansour. PAC bounds for multi-armed bandit and Markov decision processes. In *15th Annual Conference on Computational Learning Theory, COLT 2002, Sydney, Australia, July 2002, Proceedings*, volume 2375 of *Lecture Notes in Artificial Intelligence*, pages 255–270. Springer, 2002.

- [56] C. N. Fiechter. Efficient reinforcement learning. In *Proc. 7th Annu. ACM Conf. on Comput. Learning Theory*, pages 88–97. ACM Press, New York, NY, 1994.
- [57] D. J. Finton and Y. H. Hu. Importance-based feature extraction for reinforcement learning. In T. Petsche, editor, *Computational Learning Theory and Natural Learning Systems*, volume III: Selecting Good Models, chapter 5, pages 77–94. MIT Press, 1995.
- [58] J. S. Frame. Mean deviation of the binomial distribution. *The American Mathematical Monthly*, 52(7):377–379, 1945.
- [59] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139, 1997.
- [60] Z. Gábor, Z. Kalmár, and C. Szepesvári. Multi-criteria reinforcement learning. In *Proc. 15th International Conf. on Machine Learning*, pages 197–205. Morgan Kaufmann, San Francisco, CA, 1998.
- [61] L. M. Gambardella and M. Dorigo. Ant-Q: a reinforcement learning approach to the traveling salesman problem. In *Proc. 12th International Conference on Machine Learning*, pages 252–260. Morgan Kaufmann, 1995.
- [62] F. Garcia and S. M. Ndiaye. A learning rate analysis of reinforcement learning algorithms in finite-horizon. In *Proc. 15th International Conf. on Machine Learning*, pages 215–223. Morgan Kaufmann, San Francisco, CA, 1998.
- [63] P. Geibel. Reinforcement learning with bounded risk. In *Proc. 18th International Conf. on Machine Learning*, pages 162–169. Morgan Kaufmann, San Francisco, CA, 2001.
- [64] S. Geulen, B. Voeking, and M. Winkler. Regret minimization for online buffering problems using the weighted majority algorithm. In A. T. Kalai and M. Mohri, editors, *Proceedings of the 23rd Conference on Learning Theory*, pages 132–143. Omnipress, June 2010.
- [65] Z. Ghahramani and G. E. Hinton. Variational learning for switching state-space models. *Neural Computation*, 12(4):831–864, 2000.

- [66] Z. Ghahramani and M. I. Jordan. Factorial hidden markov models. *Machine Learning*, 29(2-3):245–273, 1997. ISSN 0885-6125.
- [67] M. Ghavamzadeh and S. Mahadevan. Continuous-time hierarchical reinforcement learning. In *Proc. 18th International Conf. on Machine Learning*, pages 186–193. Morgan Kaufmann, San Francisco, CA, 2001.
- [68] M. R. Glickman and K. Sycara. Evolutionary search, stochastic policies with memory, and reinforcement learning with hidden state. In *Proc. 18th International Conf. on Machine Learning*, pages 194–201. Morgan Kaufmann, San Francisco, CA, 2001.
- [69] D. E. Goldberg. Probability matching, the magnitude of reinforcement, and classifier system bidding. *Machine Learning*, 5:407–425, 1990.
- [70] R. B. Gramacy, M. K. Warmuth, S. A. Brandt, and I. Ari. Adaptive caching by refetching. In *In Advances in Neural Information Processing Systems 15*, pages 1465–1472. MIT Press, 2002.
- [71] P. D. Grünwald. *The Minimum Description Length Principle*. The MIT Press, 2007.
- [72] D. Gusfield. Connectivity and edge-disjoint spanning trees. *Information Processing Letters*, 16(2):87–89, 1983.
- [73] J. Hannan. Approximation to Bayes risk in repeated play. In *Contributions to the Theory of Games*, volume 3, pages 97–139. Princeton University Press, 1957.
- [74] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer Verlag, 2001.
- [75] D. Haussler, J. Kivinen, and M. K. Warmuth. Sequential prediction of individual sequences under general loss functions. *IEEE Transactions on Information Theory*, 44(5):1906–1925, 1998.
- [76] M. Heger. Consideration of risk in reinforcement learning. In *Proc. 11th International Conference on Machine Learning*, pages 105–111. Morgan Kaufmann, 1994.

- [77] D. P. Helmbold and M. K. Warmuth. Learning permutations with exponential weights. *Journal of Machine Learning Research*, 10:1705–1736, July 2009.
- [78] D. P. Helmbold, D. D. E. Long, T. L. Sconyers, and B. Sherrod. Adaptive disk spin-down for mobile computers. *ACM/Baltzer Mobile Networks and Applications (MONET)*, pages 285–297, 2000.
- [79] M. Herbster and M. K. Warmuth. Tracking the best expert. In *Proceedings of the 12th Annual Conference on Learning Theory (COLT 1995)*, pages 286–294, 1995.
- [80] M. Herbster and M. K. Warmuth. Tracking the best expert. *Machine Learning*, 32:151–178, 1998.
- [81] M. Herbster and M. K. Warmuth. Tracking the best linear predictor. *Journal of Machine Learning Research*, 1:281–309, 2001.
- [82] D. F. Hougen, M. Gini, and J. Slagle. An integrated connectionist approach to reinforcement learning for robotic control: The advantages of indexed partitioning. In *Proc. 17th International Conf. on Machine Learning*, pages 383–390. Morgan Kaufmann, San Francisco, CA, 2000.
- [83] J. Hu and M. P. Wellman. Multiagent reinforcement learning: theoretical framework and an algorithm. In *Proc. 15th International Conf. on Machine Learning*, pages 242–250. Morgan Kaufmann, San Francisco, CA, 1998.
- [84] M. Hutter and J. Poland. Adaptive online prediction by following the perturbed leader. *Journal of Machine Learning Research*, 6:639–660, Apr. 2005.
- [85] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *Journal of the ACM*, 51(4):671–697, 2004. ISSN 0004-5411.
- [86] L. ji Lin. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, 8:293–321, 1992.

- [87] V. Kachitvichyanukul and B. W. Schmeiser. Binomial random variate generation. *Commun. ACM*, 31(2):216–222, 1988. ISSN 0001-0782.
- [88] L. P. Kaelbling. Associative methods in reinforcement learning: an empirical study. In S. J. Hanson, T. Petsche, R. L. Rivest, and M. Kearns, editors, *Computational Learning Theory and Natural Learning Systems*, volume II: Intersections Between Theory and Experiment, chapter 9, pages 133–153. MIT Press, 1994.
- [89] L. P. Kaelbling. Associative reinforcement learning: Functions in  $k$ -DNF. *Machine Learning*, 15(3):279–298, 1994.
- [90] L. P. Kaelbling. Associative reinforcement learning: A generate and test algorithm. *Machine Learning*, 15(3):299–319, 1994.
- [91] A. Kalai. A perturbation that makes “Follow the Leader” equivalent to “Randomized Weighted Majority”. Private communication, Dec. 2005.
- [92] A. Kalai and S. Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005. ISSN 0022-0000.
- [93] M. Kearns and S. Singh. Near-optimal reinforcement learning in polynomial time. In *Proc. 15th International Conf. on Machine Learning*, pages 260–268. Morgan Kaufmann, San Francisco, CA, 1998.
- [94] M. Kearns and S. Singh. Near-optimal reinforcement learning in polynomial time. *Machine Learning*, 49(2-3):209–232, 2002.
- [95] M. J. Kearns and U. V. Vazirani. *An Introduction to Computational Learning Theory*. MIT Press, Cambridge, Massachusetts, 1994.
- [96] H. Kimura and S. Kobayashi. An analysis of actor/critic algorithms using eligibility traces: reinforcement learning with imperfect value functions. In *Proc. 15th International Conf. on Machine Learning*, pages 278–286. Morgan Kaufmann, San Francisco, CA, 1998.

- [97] H. Kimura, M. Yamamura, and S. Kobayashi. Reinforcement learning by stochastic hill climbing on discounted reward. In *Proc. 12th International Conference on Machine Learning*, pages 295–303. Morgan Kaufmann, 1995.
- [98] H. Kimura, K. Miyazaki, and S. Kobayashi. Reinforcement learning in POMDPs with function approximation. In *Proc. 14th International Conference on Machine Learning*, pages 152–160. Morgan Kaufmann, 1997.
- [99] T. Koo, A. Globerson, X. Carreras, and M. Collins. Structured prediction models via the Matrix-Tree theorem. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 141–150, 2007.
- [100] W. M. Koolen and S. de Rooij. Combining expert advice efficiently. In R. Servedio and T. Zang, editors, *Proceedings of the 21st Annual Conference on Learning Theory (COLT 2008)*, pages 275–286, June 2008.
- [101] W. M. Koolen and S. de Rooij. Combining expert advice efficiently. arXiv:0802.2015, Feb. 2008.
- [102] D. Kuzmin and M. K. Warmuth. Optimum follow the leader algorithm. In *Proceedings of the 18th Annual Conference on Learning Theory (COLT '05)*, pages 684–686. Springer-Verlag, June 2005. Open problem.
- [103] M. G. Lagoudakis and M. L. Littman. Algorithm selection using reinforcement learning. In *Proc. 17th International Conf. on Machine Learning*, pages 511–518. Morgan Kaufmann, San Francisco, CA, 2000.
- [104] N. Landwehr. Modeling interleaved hidden processes. In *Proceedings of the 25th international conference on Machine learning*, pages 520–527, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-205-4.
- [105] M. Lauer and M. Riedmiller. An algorithm for distributed reinforcement learning in cooperative multi-agent systems. In *Proc.*

- 17th International Conf. on Machine Learning*, pages 535–542. Morgan Kaufmann, San Francisco, CA, 2000.
- [106] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag New York, Inc., 1997.
- [107] L. Lin. Self-improving reactive agents: case studies of reinforcement learning frameworks. Technical Report CMU-CS-90-109, Carnegie Mellon Computer Science Department, Aug. 1990.
- [108] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994. Preliminary version appeared in the Proceedings of the 30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, North Carolina, 1989.
- [109] M. L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Proc. 11th International Conference on Machine Learning*, pages 157–163. Morgan Kaufmann, 1994.
- [110] M. L. Littman and C. Szepesvári. A generalized reinforcement-learning model: Convergence and applications. In *Proc. 13th International Conference on Machine Learning*, pages 310–318. Morgan Kaufmann, 1996.
- [111] R. Maclin and J. W. Shavlik. Creating advice-taking reinforcement learners. *Machine Learning*, 22:251–281, 1996.
- [112] T. L. Magnanti and L. A. Wolsey. Optimal trees. In M. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, editors, *Network Models*, volume 7 of *Handbooks in Operations Research and Management Science*, pages 503–615. North-Holland, 1995.
- [113] S. Mahadevan. To discount or not to discount in reinforcement learning: a case study comparing R learning and Q learning. In *Proc. 11th International Conference on Machine Learning*, pages 164–172. Morgan Kaufmann, 1994.
- [114] S. Mahadevan. Average reward reinforcement learning: Foundations, algorithms, and empirical results. *Machine Learning*, 22: 159–195, 1996.

- [115] S. Mahadevan. Sensitive discount optimality: unifying discounted and average reward reinforcement learning. In *Proc. 13th International Conference on Machine Learning*, pages 328–336. Morgan Kaufmann, 1996.
- [116] S. Mahadevan and J. Connell. Automatic programming of behavior-based robots using reinforcement learning. Technical report, IBM Research at Yorktown Heights, Dec. 1990.
- [117] S. Mannor and N. Shimkin. A geometric approach to multi-criterion reinforcement learning. *Journal of Machine Learning Research*, 5:325–360, 2004.
- [118] S. Mannor and J. N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5:623–648, 2004.
- [119] Y. Mansour. Reinforcement learning and mistake bounded algorithms. In *Proc. 12th Annu. Conf. on Comput. Learning Theory*, pages 183–192. ACM Press, New York, NY, 1999.
- [120] C. E. Mariano and E. F. Morales. DQL: A new updating strategy for reinforcement learning based on Q-learning. In *Machine Learning: ECML 2001, 12th European Conference on Machine Learning, Freiburg, Germany, September 5-7, 2001, Proceedings*, volume 2167 of *Lecture Notes in Artificial Intelligence*, pages 324–335. Springer, 2001.
- [121] R. A. McCallum. Instance-based utile distinctions for reinforcement learning with hidden state. In *Proc. 12th International Conference on Machine Learning*, pages 387–395. Morgan Kaufmann, 1995.
- [122] A. McGovern and A. G. Barto. Automatic discovery of subgoals in reinforcement learning using diverse density. In *Proc. 18th International Conf. on Machine Learning*, pages 361–368. Morgan Kaufmann, San Francisco, CA, 2001.
- [123] A. McGovern, E. Moss, and A. G. Barto. Building a basic block instruction scheduler with reinforcement learning and rollouts. *Machine Learning*, 49(2-3):141–160, 2002.

- [124] O. Mihatsch and R. Neuneier. Risk-sensitive reinforcement learning. *Machine Learning*, 49(2-3):267–290, 2002.
- [125] J. D. R. Millán and C. Torras. A reinforcement connectionist approach to robot path finding in non-maze-like environments. *Machine Learning*, 8:363–395, 1992.
- [126] M. Minsky and S. Papert. *Perceptrons*. MIT Press, Cambridge, MA, 1988.
- [127] T. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [128] A. Moffat. *Compression and Coding Algorithms*. Kluwer Academic Publishers, 2002. ISBN 0-7923-7668-4.
- [129] C. Monteleoni and T. Jaakkola. Online learning of non-stationary sequences. *Advances in Neural Information Processing Systems*, 16: 1093–1100, 2003.
- [130] A. W. Moore. Reinforcement learning in factories: the auton project (abstract). In *Proc. 13th International Conference on Machine Learning*, page 556. Morgan Kaufmann, 1996.
- [131] A. W. Moore and C. G. Atkeson. Prioritized sweeping: Reinforcement learning with less data and less time. *Machine Learning*, 13: 103–130, 1993.
- [132] D. E. Moriarty and R. Miikkulainen. Efficient reinforcement learning through symbiotic evolution. *Machine Learning*, 22:11–32, 1996.
- [133] J. Morimoto and K. Doya. Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning. In *Proc. 17th International Conf. on Machine Learning*, pages 623–630. Morgan Kaufmann, San Francisco, CA, 2000.
- [134] R. Munos. A convergent reinforcement learning algorithm in the continuous case: the finite-element reinforcement learning. In *Proc. 13th International Conference on Machine Learning*, pages 337–345. Morgan Kaufmann, 1996.

- [135] R. Munos. A study of reinforcement learning in the continuous case by the means of viscosity solutions. *Machine Learning*, 40(3): 265–299, 2000.
- [136] B. K. Natarajan and P. Tadepalli. Two new frameworks for learning. In *Proc. of the 5th International Conference on Machine Learning*, pages 402–415, San Mateo, CA, June 1988. published by Morgan Kaufmann.
- [137] A. Y. Ng and S. Russell. Algorithms for inverse reinforcement learning. In *Proc. 17th International Conf. on Machine Learning*, pages 663–670. Morgan Kaufmann, San Francisco, CA, 2000.
- [138] D. Ormoneit and Š. Sen. Kernel-based reinforcement learning. *Machine Learning*, 49(2-3):161–178, 2002.
- [139] M. D. Pendrith and M. J. McGarity. An analysis of direct reinforcement learning in non-Markovian domains. In *Proc. 15th International Conf. on Machine Learning*, pages 421–429. Morgan Kaufmann, San Francisco, CA, 1998.
- [140] M. D. Pendrith and M. R. K. Ryan. Actual return reinforcement learning versus temporal differences: some theoretical and experimental results. In *Proc. 13th International Conference on Machine Learning*, pages 373–381. Morgan Kaufmann, 1996.
- [141] T. J. Perkins and A. G. Barto. Lyapunov-constrained action sets for reinforcement learning. In *Proc. 18th International Conf. on Machine Learning*, pages 409–416. Morgan Kaufmann, San Francisco, CA, 2001.
- [142] T. J. Perkins and A. G. Barto. Lyapunov design for safe reinforcement learning. *Journal of Machine Learning Research*, 3:803–832, 2002.
- [143] J. Poland. FPL analysis for adaptive bandits. In *Stochastic Algorithms: Foundations and Applications, Third International Symposium, SAGA 2005, Moscow, Russia, October 2005, Proceedings*, volume 3777 of *Lecture Notes in Computer Science*, pages 58–69. Springer, 2005.

- [144] D. Precup and R. S. Sutton. Exponentiated gradient methods for reinforcement learning. In *Proc. 14th International Conference on Machine Learning*, pages 272–277. Morgan Kaufmann, 1997.
- [145] B. Price and C. Boutilier. Implicit imitation in multiagent reinforcement learning. In *Proc. 16th International Conf. on Machine Learning*, pages 325–334. Morgan Kaufmann, San Francisco, CA, 1999.
- [146] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77, issue 2, pages 257–285, 1989.
- [147] J. Randalø. Shaping in reinforcement learning by changing the physics of the problem. In *Proc. 17th International Conf. on Machine Learning*, pages 767–774. Morgan Kaufmann, San Francisco, CA, 2000.
- [148] J. Randalø and P. Alstrøm. Learning to drive a bicycle using reinforcement learning and shaping. In *Proc. 15th International Conf. on Machine Learning*, pages 463–471. Morgan Kaufmann, San Francisco, CA, 1998.
- [149] J. Randalø, A. G. Barto, and M. T. Rosenstein. Combining reinforcement learning with a local control algorithm. In *Proc. 17th International Conf. on Machine Learning*, pages 775–782. Morgan Kaufmann, San Francisco, CA, 2000.
- [150] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006.
- [151] J. Rennie and A. K. McCallum. Using reinforcement learning to spider the web efficiently. In *Proc. 16th International Conf. on Machine Learning*, pages 335–343. Morgan Kaufmann, San Francisco, CA, 1999.
- [152] S. I. Reynolds. Adaptive resolution model-free reinforcement learning: Decision boundary partitioning. In *Proc. 17th International Conf. on Machine Learning*, pages 783–790. Morgan Kaufmann, San Francisco, CA, 2000.

- [153] C. Richter and J. Stachowiak. Knowledge propagation in model-based reinforcement learning tasks. In *Proc. 17th International Conf. on Machine Learning*, pages 791–798. Morgan Kaufmann, San Francisco, CA, 2000.
- [154] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*, volume 15 of *Series in Computer Science*. World Scientific, 1989.
- [155] R. L. Rivest and Y. Yin. Simulation results for a new two-armed bandit heuristic. In S. J. Hanson, G. A. Drastal, and R. L. Rivest, editors, *Computational Learning Theory and Natural Learning Systems*, volume I: Constraints and Prospects, chapter 17, pages 477–486. MIT Press, 1994. Earlier version in 1990 Conference on Computation Learning and Natural Learning at Princeton.
- [156] M. Ryan and M. Reid. Learning to fly: An application of hierarchical reinforcement learning. In *Proc. 17th International Conf. on Machine Learning*, pages 807–814. Morgan Kaufmann, San Francisco, CA, 2000.
- [157] M. R. K. Ryan and M. D. Pendrith. RL-TOPs: an architecture for modularity and re-use in reinforcement learning. In *Proc. 15th International Conf. on Machine Learning*, pages 481–487. Morgan Kaufmann, San Francisco, CA, 1998.
- [158] M. Salganicoff and L. H. Ungar. Active exploration and learning in real-valued spaces using multi-armed bandit allocation indices. In *Proc. 12th International Conference on Machine Learning*, pages 480–487. Morgan Kaufmann, 1995.
- [159] M. Sato and S. Kobayashi. Average-reward reinforcement learning for variance penalized Markov decision problems. In *Proc. 18th International Conf. on Machine Learning*, pages 473–480. Morgan Kaufmann, San Francisco, CA, 2001.
- [160] B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.
- [161] A. Schrijver. *Combinatorial Optimization - Polyhedra and Efficiency*. Springer-Verlag, Berlin, 2003.

- [162] G. Shafer, A. Shen, N. Vereshchagin, and V. Vovk. Test martingales, Bayes factors, and p-values. *Statistical Science*, 2011. To appear. Preprint available as arXiv:0912.4269.
- [163] J. Shawe-Taylor, P. L. Bartlett, R. C. Williamson, and M. Anthony. A framework for structural risk minimization. In *Proc. 9th Annu. Conf. on Comput. Learning Theory*, pages 68–76. ACM Press, New York, NY, 1996.
- [164] S. P. Singh and R. S. Sutton. Reinforcement learning with replacing eligibility traces. *Machine Learning*, 22:123–158, 1996.
- [165] W. D. Smart and L. P. Kaelbling. Practical reinforcement learning in continuous spaces. In *Proc. 17th International Conf. on Machine Learning*, pages 903–910. Morgan Kaufmann, San Francisco, CA, 2000.
- [166] P. Stone and R. S. Sutton. Scaling reinforcement learning toward RoboCup soccer. In *Proc. 18th International Conf. on Machine Learning*, pages 537–544. Morgan Kaufmann, San Francisco, CA, 2001.
- [167] M. Strens. A Bayesian framework for reinforcement learning. In *Proc. 17th International Conf. on Machine Learning*, pages 943–950. Morgan Kaufmann, San Francisco, CA, 2000.
- [168] R. S. Sutton. Reinforcement learning architectures for animats. In *First International Conference on Simulation of Adaptive Behavior*, 1991.
- [169] R. S. Sutton. Open theoretical questions in reinforcement learning. In *Computational Learning Theory, 4th European Conference, EuroCOLT '99, Nordkirchen, Germany, March 29-31, 1999, Proceedings*, volume 1572 of *Lecture Notes in Artificial Intelligence*, pages 11–17. Springer, 1999.
- [170] P. Tadepalli and T. G. Dietterich. Hierarchical explanation-based reinforcement learning. In *Proc. 14th International Conference on Machine Learning*, pages 358–366. Morgan Kaufmann, 1997.
- [171] P. Tadepalli and D. Ok. Scaling up average reward reinforcement learning by approximating the domain models and the value

- function. In *Proc. 13th International Conference on Machine Learning*, pages 471–479. Morgan Kaufmann, 1996.
- [172] E. Takimoto and M. Warmuth. The last-step minimax algorithm. In *Proceedings of the 13th Annual Conference on Computational Learning Theory*, pages 100–106, 2000.
- [173] E. Takimoto and M. K. Warmuth. Path kernels and multiplicative updates. *Journal of Machine Learning Research*, 4:773–818, 2003. ISSN 1532-4435.
- [174] A. Teller and M. Veloso. Efficient learning through evolution: Neural programming and internal reinforcement. In *Proc. 17th International Conf. on Machine Learning*, pages 959–966. Morgan Kaufmann, San Francisco, CA, 2000.
- [175] H. Tong and T. X. Brown. Reinforcement learning for call admission control and routing under quality of service constraints in multimedia networks. *Machine Learning*, 49(2-3):111–139, 2002.
- [176] J. N. Tsitsiklis. A lemma on the multiarmed bandit problem. *IEEE Transactions on Automatic Control*, AC-31(6), June 1986.
- [177] T. van Erven, P. Grünwald, and S. de Rooij. Catching up faster by switching sooner: a prequential solution to the AIC-BIC dilemma. Submitted. Preprint available as arXiv:0807.1005., 2008.
- [178] T. van Erven, P. D. Grünwald, and S. de Rooij. Catching up faster in Bayesian model selection and model averaging. In *Advances in Neural Information Processing Systems 20 (NIPS 2007)*, 2008.
- [179] V. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.
- [180] P. A. Volf and F. M. Willems. Switching between two universal source coding algorithms. In *Proceedings of the Data Compression Conference, Snowbird, Utah*, pages 491–500, 1998.
- [181] V. Vovk. Aggregating strategies. In *Proceedings of the third Annual Conference on Computational Learning Theory (COLT)*, pages 371–383, 1990.
- [182] V. Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56:153–173, 1998.

- [183] V. Vovk. Derandomizing stochastic prediction strategies. *Machine Learning*, 35:247–282, 1999.
- [184] D. J. Ward and D. J. C. MacKay. Artificial intelligence: Fast hands-free writing by gaze direction. *Nature*, 418(6900):838–841, Aug. 2002.
- [185] M. K. Warmuth and D. Kuzmin. Randomized online PCA algorithms with regret bounds that are logarithmic in the dimension. *Journal of Machine Learning Research*, 9:2287–2320, Oct. 2008.
- [186] M. K. Warmuth, K. Glocer, and S. Vishwanathan. Entropy regularized LPBoost. In Y. Freund, L. Györfi, G. Turán, and T. Zeugmann, editors, *Proceedings of the 19th International Conference on Algorithmic Learning Theory (ALT '08)*, pages 256–271. Springer-Verlag, Oct. 2008.
- [187] M. Wiering. Multi-agent reinforcement learning for traffic light control. In *Proc. 17th International Conf. on Machine Learning*, pages 1151–1158. Morgan Kaufmann, San Francisco, CA, 2000.
- [188] M. A. Wiering. Reinforcement learning in dynamic environments using instantiated information. In *Proc. 18th International Conf. on Machine Learning*, pages 585–592. Morgan Kaufmann, San Francisco, CA, 2001.
- [189] F. Willems, Y. Shtarkov, and T. Tjalkens. The context tree weighting method: basic properties. *IEEE Transactions on Information Theory*, 41(3):653–664, 1995.
- [190] F. M. Willems. Coding for a binary independent piecewise-identically distributed source. *IEEE Transactions on Information Theory*, 42(6):2210–2217, Nov. 1996.
- [191] J. L. Wyatt. Exploration control in reinforcement learning using optimistic model selection. In *Proc. 18th International Conf. on Machine Learning*, pages 593–600. Morgan Kaufmann, San Francisco, CA, 2001.