Mathematics of Machine Learning:
Reinforcement learning

36 16, ~~2~~, 22, 29	1	The reinforcement learning problem: environment, agent, reward, return, Markov property, MDP, state-value V function, action Q function, policy, Bellman equation, examples (Chap. 3)
31, 37, ~~38~~, 41	2	Dynamic programming: iterative policy evaluation, proof of convergence, policy improvement theorem, policy iteration, value iteration (Chapter 4)
20, 25, ~~26~~, 24	3	Monte Carlo methods: policy evaluation, estimation of action values, Monte Carlo control, on- and off- policies, evaluating one policy while following another (Chapter 5).
5, 8, ~~12~~, 33	4	Temporal-difference (TD) learning: TD vs. Monte Carlo, convergence of TD (Bibl. Remarks 6.1-2), sarsa on-policy control, Q-learning off-policy control and a proof of its convergence (Bibl. Remarks 6.5, http://users.isr.ist.utl.pt/~mtjspaan/readingGroup/ProofQlearning.pdf), actor-critic methods (Chapter 6).
13, 32, ~~~~, 9	5	Eligibility traces: n-step TD methods, forward and backward views of TD(λ) and their equivalence (Chapter 7)
34 ~~(1)~~, 1, ~~4~~, ~~12~~, 18, ~~23~~	6	Generalization and function approximation: approximation of the V function, gradient descent, convergence for linear approximation (Tsitsiklis and Van Roy, 1997a), control with function approximation, Baird's counterexample, Tsitsiklis and Van Roy's counterexample (Chapter 8).
~~34, 35~~ 12, 3, 6, ~~7~~	7	Trust Region and Proximal Policy methods (https://arxiv.org/pdf/1502.05477.pdf, https://arxiv.org/pdf/1707.06347.pdf)
11, 28, 39	8	DQN and Double Q-learning
4, 40	9	Prioritized Experience Replay
38, 7	10	Reward shaping