Full text available at: http://dx.doi.org/10.1561/2200000068

Introduction to Multi-Armed Bandits

Full text available at: http://dx.doi.org/10.1561/2200000068

Other titles in Foundations and Trends® in Machine Learning

Computational Optimal Transport
Gabriel Peyre and Marco Cuturi
ISBN: 978-1-68083-550-2

An Introduction to Deep Reinforcement Learning
Vincent Francois-Lavet, Peter Henderson, Riashat Islam, Marc G. Bellemare and Joelle Pineau
ISBN: 978-1-68083-538-0

An Introduction to Wishart Matrix Moments
Adrian N. Bishop, Pierre Del Moral and Angele Niclas
ISBN: 978-1-68083-506-9

A Tutorial on Thompson Sampling
Daniel J. Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband and Zheng Wen
ISBN: 978-1-68083-470-3

Full text available at: http://dx.doi.org/10.1561/2200000068

Introduction to Multi-Armed Bandits

Aleksandrs Slivkins
Microsoft Research NYC
slivkins@microsoft.com

now
the essence of knowledge
Boston — Delft

Full text available at: http://dx.doi.org/10.1561/2200000068

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

A. Slivkins. *Introduction to Multi-Armed Bandits*. Foundations and Trends® in Machine Learning, vol. 12, no. 1-2, pp. 1–286, 2019.

ISBN: 978-1-68083-621-9
© 2019 A. Slivkins

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Full text available at: http://dx.doi.org/10.1561/2200000068

Foundations and Trends® in Machine Learning
Volume 12, Issue 1-2, 2019

Editorial Board

Editor-in-Chief

Michael Jordan
University of California, Berkeley
United States

Editors

Peter Bartlett
UC Berkeley

Yoshua Bengio
Université de Montréal

Avrim Blum
Toyota Technological Institute

Craig Boutilier
University of Toronto

Stephen Boyd
Stanford University

Carla Brodley
Northeastern University

Inderjit Dhillon
Texas at Austin

Jerome Friedman
Stanford University

Kenji Fukumizu
ISM

Zoubin Ghahramani
Cambridge University

David Heckerman
Amazon

Tom Heskes
Radboud University

Geoffrey Hinton
University of Toronto

Aapo Hyvarinen
Helsinki IIT

Leslie Pack Kaelbling
MIT

Michael Kearns
UPenn

Daphne Koller
Stanford University

John Lafferty
Yale

Michael Littman
Brown University

Gabor Lugosi
Pompeu Fabra

David Madigan
Columbia University

Pascal Massart
Université de Paris-Sud

Andrew McCallum
University of Massachusetts Amherst

Marina Meila
University of Washington

Andrew Moore
CMU

John Platt
Microsoft Research

Luc de Raedt
KU Leuven

Christian Robert
Paris-Dauphine

Sunita Sarawagi
IIT Bombay

Robert Schapire
Microsoft Research

Bernhard Schoelkopf
Max Planck Institute

Richard Sutton
University of Alberta

Larry Wasserman
CMU

Bin Yu
UC Berkeley

Full text available at: http://dx.doi.org/10.1561/2200000068

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, 2019, Volume 12, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Full text available at: http://dx.doi.org/10.1561/2200000068

Contents

Full text available at: http://dx.doi.org/10.1561/2200000068

Full text available at: http://dx.doi.org/10.1561/2200000068

Full text available at: http://dx.doi.org/10.1561/2200000068

Full text available at: http://dx.doi.org/10.1561/2200000068

Introduction to Multi-Armed Bandits

Aleksandrs Slivkins

Microsoft Research NYC; slivkins@microsoft.com

ABSTRACT

Multi-armed bandits a simple but very powerful framework for algorithms that make decisions over time under uncertainty. An enormous body of work has accumulated over the years, covered in several books and surveys. This book provides a more introductory, textbook-like treatment of the subject. Each chapter tackles a particular line of work, providing a self-contained, teachable technical introduction and a brief review of the further developments.

Aleksandrs Slivkins (2019), "Introduction to Multi-Armed Bandits", Foundations and Trends® in Machine Learning: Vol. 12, No. 1-2, pp 1–286. DOI: 10.1561/2200000068.

Full text available at: http://dx.doi.org/10.1561/2200000068

Preface

Multi-armed bandits is a rich, multi-disciplinary area studied since Thompson (1933), with a big surge of activity in the past 10-15 years. An enormous body of work has accumulated over the years. While various subsets of this work have been covered in depth in several books and surveys (Berry and Fristedt, 1985; Cesa-Bianchi and Lugosi, 2006; Bergemann and Välimäki, 2006; Gittins *et al.*, 2011; Bubeck and Cesa-Bianchi, 2012), this book provides a more textbook-like treatment of the subject.

The organizing principles for this book can be summarized as follows. The work on multi-armed bandits can be partitioned into a dozen or so lines of work. Each chapter tackles one line of work, providing a self-contained introduction and pointers for further reading. We favor fundamental ideas and elementary, teachable proofs over the strongest possible results. We emphasize accessibility of the material: while exposure to machine learning and probability/statistics would certainly help, a standard undergraduate course on algorithms, *e.g.,* one based on Kleinberg and Tardos (2005), should suffice for background. With the above principles in mind, the choice specific topics and results is based on the author's subjective understanding of what is important

Full text available at: http://dx.doi.org/10.1561/2200000068

and "teachable" (*i.e.,* presentable in a relatively simple manner). Many important results has been deemed too technical or advanced to be presented in detail.

This book is based on a graduate course at University of Maryland, College Park, taught by the author in Fall 2016. Each chapter corresponds to a week of the course, and is based on the lecture notes. Five chapters were used in a similar course at Columbia University, co-taught by the author in Fall 2017.

To keep the book manageable, and also more accessible, we chose not to dwell on the deep connections to online convex optimization. A modern treatment of this fascinating subject can be found, *e.g.,* in the recent textbook by Hazan (2015). Likewise, we chose not venture into a much more general problem space of reinforcement learning, a subject of many graduate courses and textbooks such as Sutton and Barto (1998) and Szepesvári (2010). A course based on this book would be complementary to graduate-level courses on online convex optimization and reinforcement learning. Also, we do not discuss MDP-based models of multi-armed bandits and the Gittins algorithm; this direction is covered in Gittins *et al.* (2011).

The book is structured as follows. The first four chapters are on IID rewards, from the basic model to impossibility results to Bayesian priors to Lipschitz rewards. The next three chapters are on adversarial rewards, from the full-feedback version to adversarial bandits to extensions with linear rewards and combinatorially structured actions. Chapter 8 is on contextual bandits, a middle ground between IID and adversarial bandits in which the change in reward distributions is completely explained by observable contexts. The remaining chapters cover connections to economics, from learning in repeated games to a (generalization of) dynamic pricing with limited supply to exploration in the presence of incentives. Each chapter contains a section on bibliographic notes and further directions. Many of the chapters conclude with some exercises. Appendix A provides a self-sufficient background on concentration inequalities.

On a final note, the author encourages colleagues to use this book in their courses. A brief email regarding which chapters have been used, along with any feedback, would be appreciated.

Full text available at: http://dx.doi.org/10.1561/2200000068

An excellent book on multi-armed bandits, Lattimore and Szepesvári (2019), will appear later this year. This book is much larger than ours; it provides a deeper treatment for a number of topics, and omits a few others. Evolving simultaneously and independently over the past 2-3 years, our books reflect the authors' somewhat differing tastes and presentation styles, and, I believe, are complementary to one another.

Full text available at: http://dx.doi.org/10.1561/2200000068

Introduction: Scope and Motivation

Multi-armed bandits is a simple but very powerful framework for algorithms that make decisions over time under uncertainty. Let us outline some of the problems that fall under this framework.

We start with three running examples, concrete albeit very stylized:

News website When a new user arrives, a website site picks an article header to show, observes whether the user clicks on this header. The site's goal is maximize the total number of clicks.

Dynamic pricing A store is selling a digital good, *e.g.,* an app or a song. When a new customer arrives, the store chooses a price offered to this customer. The customer buys (or not) and leaves forever. The store's goal is to maximize the total profit.

Investment Each morning, you choose one stock to invest into, and invest \$1. In the end of the day, you observe the change in value for each stock. The goal is to maximize the total wealth.

Multi-armed bandits unifies these examples (and many others). In the basic version, an algorithm has K possible actions to choose from, a.k.a. *arms*, and T rounds. In each round, the algorithm chooses an arm and

Full text available at: http://dx.doi.org/10.1561/2200000068

collects a reward for this arm. The reward is drawn independently from some distribution which is fixed (*i.e.,* depends only on the chosen arm), but not known to the algorithm. Going back to the running examples:

Example	Action	Reward
News website	an article to display	1 if clicked, 0 otherwise
Dynamic pricing	a price to offer	p if sale, 0 otherwise
Investment	a stock to invest into	change in value

In the basic model, an algorithm observes the reward for the chosen arm after each round, but not for the other arms that could have been chosen. Therefore, the algorithm typically needs to *explore*: try out different arms to acquire new information. Indeed, if an algorithm always chooses arm 1, how would it know if arm 2 is better? Thus, we have a tradeoff between exploration and *exploitation*: making optimal near-term decisions based on the available information. This tradeoff, which arises in numerous application scenarios, is essential in multi-armed bandits. Essentially, the algorithm strives to learn which arms are best (perhaps approximately so), while not spending too much time exploring.

The term "multi-armed bandits" comes from a stylized gambling scenario in which a gambler faces several slot machines, a.k.a. one-armed bandits, that appear identical, but yield different payoffs.

Multi-dimensional problem space

Multi-armed bandits is a huge problem space, with many "dimensions" along which the models can be made more expressive and closer to reality. We discuss some of these modeling dimensions below. Each dimension gave rise to a prominent line of work, discussed later in this book.

Auxiliary feedback. What feedback is available to the algorithm after each round, other than the reward for the chosen arm? Does the algorithm observe rewards for the other arms? Let's check our examples:

Full text available at: http://dx.doi.org/10.1561/2200000068

Example	Auxiliary feedback	Rewards for any other arms?
News website	N/A	no (*bandit feedback*).
Dynamic pricing	sale ⇒ sale at any lower price, no sale ⇒ no sale at any higher price	yes, for some arms, but not for all arms (*partial feedback*).
Investment	change in value for all other stocks	yes, for all arms (*full feedback*).

We distinguish three types of feedback: *bandit feedback*, when the algorithm observes the reward for the chosen arm, and no other feedback; *full feedback*, when the algorithm observes the rewards for all arms that could have been chosen; and *partial feedback*, when some information is revealed, in addition to the reward of the chosen arm, but it does not always amount to full feedback.

This book mainly focuses on problems with bandit feedback. We also cover some of the fundamental results on full feedback, which are essential for developing subsequent bandit results. Partial feedback sometimes arises in extensions and special cases, and can be used to improve performance.

Rewards model. Where do the rewards come from? Several alternatives has been studied:

- *IID rewards:* the reward for each arm is drawn independently from a fixed distribution that depends on the arm but not on the round t.
- *Adversarial rewards:* rewards can be arbitrary, as if they are chosen by an "adversary" that tries to fool the algorithm.
- *Constrained adversary:* rewards are chosen by an adversary that is subject to some constraints, *e.g.,* reward of each arm cannot change much from one round to another, or the reward of each arm can change at most a few times, or the total change in rewards is upper-bounded.
- *Stochastic rewards* (beyond IID): rewards evolves over time as a random process, *e.g.,* a random walk.

Full text available at: http://dx.doi.org/10.1561/2200000068

Contexts. In each round, an algorithm may observe some *context* before choosing an action. Such context often comprises the known properties of the current user, and allows for personalized actions.

Example	Context
News website	user location and demographics
Dynamic pricing	customer's device, location, demographics
Investment	current state of the economy.

The algorithm has a different high-level objective. It is no longer interested in learning one good arm, since any one arm may be great for some contexts, and terrible for some others. Instead, it strives to learn the best *policy* which maps contexts to arms (while not spending too much time exploring).

Bayesian priors. Each problem can be studied under a *Bayesian* approach, whereby the problem instance comes from a known distribution (called *Bayesian prior*). One is typically interested in provable guarantees in expectation over this distribution.

Structured rewards. Rewards may have a known structure, *e.g.,* arms correspond to points in $\mathbb{R}^d$, and in each round the reward is a linear (resp., concave or Lipschitz) function of the chosen arm.

Global constraints. The algorithm can be subject to global constraints that bind across arms and across rounds. For example, in dynamic pricing there may be a limited inventory of items for sale.

Structured actions. An algorithm may need to make several decisions at once, *e.g.,* a news website may need to pick a slate of articles, and a seller may need to choose prices for the entire slate of offerings.

Application domains

Multi-armed bandit problems arise in a variety of application domains. The original application has been the design of "ethical" medical trials, so as to attain useful scientific data while minimizing harm to the patients. Prominent modern applications concern the Web: from tuning the look and feel of a website, to choosing which content to highlight, to

Full text available at: http://dx.doi.org/10.1561/2200000068

optimizing web search results, to placing ads on webpages. Recommender systems can use exploration to improve its recommendations for movies, restaurants, hotels, and so forth. Another cluster of applications pertains to economics: a seller can optimize its prices and offerings; likewise, a frequent buyer such as a procurement agency can optimize its bids; an auctioneer can adjust its auction over time; a crowdsourcing platform can improve the assignment of tasks, workers and prices. In computer systems, one can experiment and learn, rather than rely on a rigid design, so as to optimize datacenters and networking protocols. Finally, one can teach a robot to better perform its tasks.

Application	Action (e.g.)	Reward (e.g.)
medical trials	which drug to prescribe	healthy/not.
web design	font color or page layout	#clicks.
web content	items/articles to emphasize	#clicks.
web search	search results given a query	#happy users.
advertisement	which ad to display	ad revenue.
recommender systems	which movie to watch	#recommendations followed.
sales	which products to offer at which prices	revenue.
procurement	which items to buy at which prices	#items procured.
auctions	which reserve price to use	revenue
crowdsourcing	which tasks to give to which workers, and at which prices	#completed tasks.
datacenters	server to route the job to	completion time.
Internet	which TCP settings to use	connection quality.
smart radios	radio frequency to use	#transmitted messages.
robot control	a “strategy” for a given task	completion time.

Full text available at: http://dx.doi.org/10.1561/2200000068

(Brief) bibliographic notes

Medical trials has a major motivation for introducing multi-armed bandits and exploration-exploitation tradeoff (Thompson, 1933; Gittins, 1979). Bandit-like designs for medical trials belong to the realm of *adaptive* medical trials (Chow and Chang, 2008), which can also include other "adaptive" features such as early stopping, sample size re-estimation, and changing the dosage.

Applications to the Web trace back to Pandey *et al.* (2007a), Pandey *et al.* (2007b), and Langford and Zhang (2007) for ad placement, Li *et al.* (2010) and Li *et al.* (2011) for news optimization, and Radlinski *et al.* (2008) for web search. A survey of the more recent literature in this direction is beyond our scope.

Bandit algorithms tailored to recommendation systems are studied, *e.g.,* in Bresler *et al.* (2014), Li *et al.* (2016), and Bresler *et al.* (2016).

Applications to problems in economics comprise many aspects: optimizing seller's prices, a.k.a. *dynamic pricing* (Boer, 2015, a survey); optimizing seller's product offerings, a.k.a. *dynamic assortment* (*e.g.,* Sauré and Zeevi, 2013; Agrawal *et al.*, 2016a); optimizing buyers prices, a.k.a. *dynamic procurement* (*e.g.,* Badanidiyuru *et al.*, 2012; Badanidiyuru *et al.*, 2018); design of auctions (*e.g.,* Bergemann and Said, 2011; Cesa-Bianchi *et al.*, 2013; Babaioff *et al.*, 2015b); design of incentives and information structures (Slivkins, 2017, a survey); design of crowdsourcing platforms (Slivkins and Vaughan, 2013, a survey).

A growing line of work on applications to Internet routing and congestion control includes Dong *et al.* (2015), Dong *et al.* (2018), Jiang *et al.* (2016), and Jiang *et al.* (2017). Early theoretical work on bandits with the same motivation is in Awerbuch and Kleinberg (2008) and Awerbuch *et al.* (2005). Bandit problems directly motivated by radio networks have been studied starting from Lai *et al.* (2008), Liu and Zhao (2010), and Anandkumar *et al.* (2011).

Full text available at: http://dx.doi.org/10.1561/2200000068

References

Abbasi-Yadkori, Y., D. Pál, and C. Szepesvári. 2011. "Improved Algorithms for Linear Stochastic Bandits". In: *25th Advances in Neural Information Processing Systems (NIPS)*. 2312–2320.

Abernethy, J. D. and J.-K. Wang. 2017. "On Frank-Wolfe and Equilibrium Computation". In: *Advances in Neural Information Processing Systems (NIPS)*. 6584–6593.

Abraham, I. and D. Malkhi. 2005. "Name independent routing for growth bounded networks". In: *17th ACM Symp. on Parallel Algorithms and Architectures (SPAA)*. 49–55.

Agarwal, A., A. Beygelzimer, M. Dudik, J. Langford, and H. Wallach. 2017a. "A reductions approach to fair classification". *Fairness, Accountability, and Transparency in Machine Learning (FATML)*.

Agarwal, A., S. Bird, M. Cozowicz, M. Dudik, L. Hoang, J. Langford, L. Li, D. Melamed, G. Oshri, S. Sen, and A. Slivkins. 2016. "Multiworld Testing: A System for Experimentation, Learning, And Decision-Making". A white paper, available at https://github.com/Microsoft/mwt-ds/raw/master/images/MWT-WhitePaper.pdf.

Agarwal, A., S. Bird, M. Cozowicz, L. Hoang, J. Langford, S. Lee, J. Li, D. Melamed, G. Oshri, O. Ribas, S. Sen, and A. Slivkins. 2017b. "Making Contextual Decisions with Low Technical Debt". Techical report at arxiv.org/abs/1606.03966.

Full text available at: http://dx.doi.org/10.1561/2200000068

Agarwal, A., M. Dudik, S. Kale, J. Langford, and R. E. Schapire. 2012. "Contextual Bandit Learning with Predictable Rewards". In: *15th Intl. Conf. on Artificial Intelligence and Statistics (AISTATS)*. 19–26.

Agarwal, A., D. Hsu, S. Kale, J. Langford, L. Li, and R. Schapire. 2014. "Taming the Monster: A Fast and Simple Algorithm for Contextual Bandits". In: *31st Intl. Conf. on Machine Learning (ICML)*.

Agarwal, A., H. Luo, B. Neyshabur, and R. E. Schapire. 2017c. "Corralling a Band of Bandit Algorithms". In: *30th Conf. on Learning Theory (COLT)*. 12–38.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt. 2005. "Competition and Innovation: An Inverted U Relationship". *Quaterly J. of Economics*. 120(2): 701–728.

Agrawal, R. 1995. "The continuum-armed bandit problem". *SIAM J. Control and Optimization*. 33(6): 1926–1951.

Agrawal, S., V. Avadhanula, V. Goyal, and A. Zeevi. 2016a. "A Near-Optimal Exploration-Exploitation Approach for Assortment Selection". In: *17th ACM Conf. on Economics and Computation (ACM EC)*. 599–600.

Agrawal, S. and N. R. Devanur. 2014. "Bandits with concave rewards and convex knapsacks". In: *15th ACM Conf. on Economics and Computation (ACM EC)*.

Agrawal, S. and N. R. Devanur. 2016. "Linear Contextual Bandits with Knapsacks". In: *29th Advances in Neural Information Processing Systems (NIPS)*.

Agrawal, S., N. R. Devanur, and L. Li. 2016b. "An efficient algorithm for contextual bandits with knapsacks, and an extension to concave objectives". In: *29th Conf. on Learning Theory (COLT)*.

Agrawal, S. and N. Goyal. 2012. "Analysis of Thompson Sampling for the multi-armed bandit problem". In: *25nd Conf. on Learning Theory (COLT)*.

Agrawal, S. and N. Goyal. 2013. "Further Optimal Regret Bounds for Thompson Sampling". In: *16th Intl. Conf. on Artificial Intelligence and Statistics (AISTATS)*. 99–107.

Full text available at: http://dx.doi.org/10.1561/2200000068

Agrawal, S., Z. Wang, and Y. Ye. 2014. "A Dynamic Near-Optimal Algorithm for Online Linear Programming". *Operations Research.* 62(4): 876–890.

Ailon, N., Z. Karnin, and T. Joachims. 2014. "Reducing Dueling Bandits to Cardinal Bandits". In: *Intl. Conf. on Machine Learning (ICML).* 856–864.

Alon, N., N. Cesa-Bianchi, O. Dekel, and T. Koren. 2015. "Online Learning with Feedback Graphs: Beyond Bandits". In: *28th Conf. on Learning Theory (COLT).* 23–35.

Alon, N., N. Cesa-Bianchi, C. Gentile, and Y. Mansour. 2013. "From Bandits to Experts: A Tale of Domination and Independence". In: *27th Advances in Neural Information Processing Systems (NIPS).* 1610–1618.

Amin, K., M. Kearns, and U. Syed. 2011. "Bandits, Query Learning, and the Haystack Dimension". In: *24th Conf. on Learning Theory (COLT).*

Amin, K., A. Rostamizadeh, and U. Syed. 2013. "Learning Prices for Repeated Auctions with Strategic Buyers". In: *26th Advances in Neural Information Processing Systems (NIPS).* 1169–1177.

Amin, K., A. Rostamizadeh, and U. Syed. 2014. "Repeated Contextual Auctions with Strategic Buyers". In: *27th Advances in Neural Information Processing Systems (NIPS).* 622–630.

Anandkumar, A., N. Michael, A. K. Tang, and A. Swami. 2011. "Distributed Algorithms for Learning and Cognitive Medium Access with Logarithmic Regret". *IEEE Journal on Selected Areas in Communications.* 29(4): 731–745.

Antos, A., G. Bartók, D. Pál, and C. Szepesvári. 2013. "Toward a classification of finite partial-monitoring games". *Theor. Comput. Sci.* 473: 77–99.

Aridor, G., A. Slivkins, and S. Wu. 2019. "The Perils of Exploration under Competition: A Computational Modeling Approach". In: *20th ACM Conf. on Economics and Computation (ACM EC).*

Arora, S., E. Hazan, and S. Kale. 2012. "The Multiplicative Weights Update Method: a Meta-Algorithm and Applications". *Theory of Computing.* 8(1): 121–164.

Full text available at: http://dx.doi.org/10.1561/2200000068

Athey, S. and I. Segal. 2013. "An Efficient Dynamic Mechanism". *Econometrica*. 81(6): 2463–2485. A preliminary version has been available as a working paper since 2007.

Audibert, J.-Y., R. Munos, and C. Szepesvári. 2009. "Exploration-exploitation Trade-off using Variance Estimates in Multi-Armed Bandits". *Theoretical Computer Science*. 410: 1876–1902.

Audibert, J. and S. Bubeck. 2010. "Regret Bounds and Minimax Policies under Partial Monitoring". *J. of Machine Learning Research (JMLR)*. 11: 2785–2836. Preliminary version in *COLT 2009*.

Audibert, J., S. Bubeck, and R. Munos. 2010. "Best Arm Identification in Multi-Armed Bandits". In: *23rd Conf. on Learning Theory (COLT)*. 41–53.

Auer, P., N. Cesa-Bianchi, and P. Fischer. 2002a. "Finite-time Analysis of the Multiarmed Bandit Problem." *Machine Learning*. 47(2-3): 235–256.

Auer, P., N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. 2002b. "The Nonstochastic Multiarmed Bandit Problem." *SIAM J. Comput.* 32(1): 48–77. Preliminary version in *36th IEEE FOCS*, 1995.

Auer, P. and C. Chiang. 2016. "An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits". In: *29th Conf. on Learning Theory (COLT)*.

Auer, P., P. Gajane, and R. Ortner. 2019. "Adaptively tracking the best arm with an unknown number of distribution changes". In: *Conf. on Learning Theory (COLT)*.

Auer, P., R. Ortner, and C. Szepesvári. 2007. "Improved Rates for the Stochastic Continuum-Armed Bandit Problem". In: *20th Conf. on Learning Theory (COLT)*. 454–468.

Aumann, R. J. 1974. "Subjectivity and correlation in randomized strategies". *J. of Mathematical Economics*. 1: 67–96.

Avner, O. and S. Mannor. 2014. "Concurrent Bandits and Cognitive Radio Networks". In: *European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*. 66–81.

Awerbuch, B., D. Holmer, H. Rubens, and R. D. Kleinberg. 2005. "Provably competitive adaptive routing". In: *24th Conf. of the IEEE Communications Society (INFOCOM)*. 631–641.

Full text available at: http://dx.doi.org/10.1561/2200000068

Awerbuch, B. and R. Kleinberg. 2008. "Online linear optimization and adaptive routing". *J. of Computer and System Sciences*. 74(1): 97–114. Preliminary version in *36th ACM STOC*, 2004.

Azar, M. G., A. Lazaric, and E. Brunskill. 2014. "Online Stochastic Optimization under Correlated Bandit Feedback". In: *31th Intl. Conf. on Machine Learning (ICML)*. 1557–1565.

Babaioff, M., S. Dughmi, R. D. Kleinberg, and A. Slivkins. 2015a. "Dynamic Pricing with Limited Supply". *ACM Trans. on Economics and Computation*. 3(1): 4. Special issue for *13th ACM EC*, 2012.

Babaioff, M., R. Kleinberg, and A. Slivkins. 2010. "Truthful Mechanisms with Implicit Payment Computation". In: *11th ACM Conf. on Electronic Commerce (EC)*. 43–52.

Babaioff, M., R. Kleinberg, and A. Slivkins. 2013. "Multi-parameter mechanisms with implicit payment computation". In: *13th ACM Conf. on Electronic Commerce (EC)*. 35–52.

Babaioff, M., R. Kleinberg, and A. Slivkins. 2015b. "Truthful Mechanisms with Implicit Payment Computation". *J. of the ACM*. 62(2): 10. Subsumes the conference papers in *ACM EC 2010* and *ACM EC 2013*.

Babaioff, M., Y. Sharma, and A. Slivkins. 2014. "Characterizing Truthful Multi-armed Bandit Mechanisms". *SIAM J. on Computing (SICOMP)*. 43(1): 194–230. Preliminary version in *10th ACM EC*, 2009.

Badanidiyuru, A., R. Kleinberg, and Y. Singer. 2012. "Learning on a budget: posted price mechanisms for online procurement". In: *13th ACM Conf. on Electronic Commerce (EC)*. 128–145.

Badanidiyuru, A., R. Kleinberg, and A. Slivkins. 2013. "Bandits with Knapsacks". In: *54th IEEE Symp. on Foundations of Computer Science (FOCS)*.

Badanidiyuru, A., R. Kleinberg, and A. Slivkins. 2018. "Bandits with Knapsacks". *J. of the ACM*. 65(3). Preliminary version in *FOCS 2013*.

Badanidiyuru, A., J. Langford, and A. Slivkins. 2014. "Resourceful Contextual Bandits". In: *27th Conf. on Learning Theory (COLT)*.

Full text available at: http://dx.doi.org/10.1561/2200000068

Bahar, G., O. Ben-Porat, K. Leyton-Brown, and M. Tennenholtz. 2019a. "Fiduciary Bandits". *CoRR*. abs/1905.07043. arXiv: 1905.07043. URL: http://arxiv.org/abs/1905.07043.

Bahar, G., R. Smorodinsky, and M. Tennenholtz. 2016. "Economic Recommendation Systems". In: *16th ACM Conf. on Electronic Commerce (EC)*.

Bahar, G., R. Smorodinsky, and M. Tennenholtz. 2019b. "Social Learning and the Innkeeper's Challenge". In: *ACM Conf. on Economics and Computation (ACM EC)*. 153–170.

Bailey, J. P. and G. Piliouras. 2018. "Multiplicative Weights Update in Zero-Sum Games". In: *ACM Conf. on Economics and Computation (ACM EC)*. 321–338.

Bartók, G., D. P. Foster, D. Pál, A. Rakhlin, and C. Szepesvári. 2014. "Partial Monitoring - Classification, Regret Bounds, and Algorithms". *Math. Oper. Res.* 39(4): 967–997.

Bastani, H., M. Bayati, and K. Khosravi. 2018. "Mostly Exploration-Free Algorithms for Contextual Bandits". *CoRR*. Working paper. arXiv: 1704.09011.

Bergemann, D. and S. Morris. 2013. "Robust Predictions in Games With Incomplete Information". *Econometrica*. 81(4): 1251–1308.

Bergemann, D. and S. Morris. 2016. "Information Design, Bayesian Persuasion and Bayes Correlated Equilibrium". Working paper.

Bergemann, D. and M. Said. 2011. "Dynamic Auctions: A Survey". In: *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons.

Bergemann, D. and J. Välimäki. 2006. "Bandit Problems". In: *The New Palgrave Dictionary of Economics, 2nd ed.* Ed. by S. Durlauf and L. Blume. Macmillan Press.

Bergemann, D. and J. Välimäki. 2010. "The Dynamic Pivot Mechanism". *Econometrica*. 78(2): 771–789. Preliminary versions have been available since 2006.

Berry, D. and B. Fristedt. 1985. *Bandit problems: sequential allocation of experiments*. Chapman&Hall.

Besbes, O. and A. Zeevi. 2009. "Dynamic Pricing Without Knowing the Demand Function: Risk Bounds and Near-Optimal Algorithms". *Operations Research*. 57(6): 1407–1420.

Full text available at: http://dx.doi.org/10.1561/2200000068

Besbes, O. and A. J. Zeevi. 2012. "Blind Network Revenue Management". *Operations Research*. 60(6): 1537–1550.

Beygelzimer, A., J. Langford, L. Li, L. Reyzin, and R. E. Schapire. 2011. "Contextual Bandit Algorithms with Supervised Learning Guarantees". In: *14th Intl. Conf. on Artificial Intelligence and Statistics (AISTATS)*.

Bietti, A., A. Agarwal, and J. Langford. 2018. "A Contextual Bandit Bake-off". *CoRR*. arXiv: 1802.04064.

Bimpikis, K., Y. Papanastasiou, and N. Savva. 2018. "Crowdsourcing Exploration". *Management Science*. 64(4): 1477–1973.

Blum, A. 1997. "Empirical support for winnow and weighted-majority based algorithms: Results on a calendar scheduling domain". *Machine Learning*. 26: 5–23.

Blum, A., M. Hajiaghayi, K. Ligett, and A. Roth. 2008. "Regret minimization and the price of total anarchy". In: *40th ACM Symp. on Theory of Computing (STOC)*. 373–382.

Blum, A., V. Kumar, A. Rudra, and F. Wu. 2003. "Online learning in online auctions". In: *14th ACM-SIAM Symp. on Discrete Algorithms (SODA)*. 202–204.

Blum, A. and Y. Mansour. 2007. "From external to internal regret". *J. of Machine Learning Research (JMLR)*. 8(13): 1307–1324. Preliminary version in *COLT 2005*.

Boer, A. V. D. 2015. "Dynamic pricing and learning: Historical origins, current research, and new directions". *Surveys in Operations Research and Management Science*. 20(1).

Bolton, P. and C. Harris. 1999. "Strategic Experimentation". *Econometrica*. 67(2): 349–374.

Boursier, E. and V. Perchet. 2018. "SIC-MMAB: Synchronisation Involves Communication in Multiplayer Multi-Armed Bandits". *CoRR*. abs/1809.08151. URL: http://arxiv.org/abs/1809.08151.

Braverman, M., J. Mao, J. Schneider, and M. Weinberg. 2018. "Selling to a No-Regret Buyer". In: *ACM Conf. on Economics and Computation (ACM EC)*. 523–538.

Braverman, M., J. Mao, J. Schneider, and S. M. Weinberg. 2019. "Multi-armed Bandit Problems with Strategic Arms". In: *Conf. on Learning Theory (COLT)*. 383–416.

Full text available at: http://dx.doi.org/10.1561/2200000068

Bresler, G., G. H. Chen, and D. Shah. 2014. "A Latent Source Model for Online Collaborative Filtering". In: *27th Advances in Neural Information Processing Systems (NIPS)*. 3347–3355.

Bresler, G., D. Shah, and L. F. Voloch. 2016. "Collaborative Filtering with Low Regret". In: *The Intl. Conf. on Measurement and Modeling of Computer Systems (SIGMETRICS)*. 207–220.

Brown, G. W. 1949. "Some notes on computation of games solutions". *Tech. rep.* No. P-78. The Rand Corporation.

Bubeck, S. 2010. "Bandits Games and Clustering Foundations". *PhD thesis*. Univ. Lille 1.

Bubeck, S. and N. Cesa-Bianchi. 2012. "Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems". *Foundations and Trends in Machine Learning*. 5(1).

Bubeck, S., M. B. Cohen, and Y. Li. 2018. "Sparsity, variance and curvature in multi-armed bandits". In: *29nd Intl. Conf. on Algorithmic Learning Theory (ALT)*.

Bubeck, S., O. Dekel, T. Koren, and Y. Peres. 2015. "Bandit Convex Optimization: $\sqrt{T}$ Regret in One Dimension". In: *28th Conf. on Learning Theory (COLT)*. 266–278.

Bubeck, S., Y. T. Lee, and R. Eldan. 2017. "Kernel-based methods for bandit convex optimization". In: *49th ACM Symp. on Theory of Computing (STOC)*. ACM. 72–85.

Bubeck, S., Y. Li, H. Luo, and C. Wei. 2019a. "Improved Path-length Regret Bounds for Bandits". In: *Conf. on Learning Theory (COLT)*.

Bubeck, S., Y. Li, Y. Peres, and M. Sellke. 2019b. "Non-Stochastic Multi-Player Multi-Armed Bandits: Optimal Rate With Collision Information, Sublinear Without". *CoRR*. abs/1904.12233. arXiv: 1904.12233. URL: http://arxiv.org/abs/1904.12233.

Bubeck, S., R. Munos, and G. Stoltz. 2011a. "Pure Exploration in Multi-Armed Bandit Problems". *Theoretical Computer Science*. 412(19): 1832–1852.

Bubeck, S., R. Munos, G. Stoltz, and C. Szepesvari. 2011b. "Online Optimization in X-Armed Bandits". *J. of Machine Learning Research (JMLR)*. 12: 1587–1627.

Full text available at: http://dx.doi.org/10.1561/2200000068

Bubeck, S. and A. Slivkins. 2012. "The best of both worlds: stochastic and adversarial bandits". In: *25th Conf. on Learning Theory (COLT)*.

Bubeck, S., G. Stoltz, and J. Y. Yu. 2011c. "Lipschitz Bandits without the Lipschitz Constant". In: *22nd Intl. Conf. on Algorithmic Learning Theory (ALT)*. 144–158.

Bull, A. 2015. "Adaptive-treed bandits". *Bernoulli J. of Statistics*. 21(4): 2289–2307.

Carpentier, A. and A. Locatelli. 2016. "Tight (Lower) Bounds for the Fixed Budget Best Arm Identification Bandit Problem". In: *29th Conf. on Learning Theory (COLT)*. 590–604.

Cesa-Bianchi, N., P. Gaillard, C. Gentile, and S. Gerchinovitz. 2017. "Algorithmic Chaining and the Role of Partial Feedback in Online Nonparametric Learning". In: *30th Conf. on Learning Theory (COLT)*. 465–481.

Cesa-Bianchi, N., C. Gentile, and Y. Mansour. 2013. "Regret Minimization for Reserve Prices in Second-Price Auctions". In: *ACM-SIAM Symp. on Discrete Algorithms (SODA)*.

Cesa-Bianchi, N. and G. Lugosi. 2003. "Potential-Based Algorithms in On-Line Prediction and Game Theory". *Machine Learning*. 51(3): 239–261.

Cesa-Bianchi, N. and G. Lugosi. 2006. *Prediction, learning, and games*. Cambridge Univ. Press.

Cesa-Bianchi, N. and G. Lugosi. 2012. "Combinatorial bandits". *J. Comput. Syst. Sci.* 78(5): 1404–1422. Preliminary version in *COLT 2009*.

Chakrabarti, D., R. Kumar, F. Radlinski, and E. Upfal. 2008. "Mortal Multi-Armed Bandits". In: *22nd Advances in Neural Information Processing Systems (NIPS)*. 273–280.

Che, Y.-K. and J. Hörner. 2018. "Optimal design for social learning". *Quarterly Journal of Economics*. Forthcoming. First published draft: 2013.

Chen, B., P. I. Frazier, and D. Kempe. 2018. "Incentivizing Exploration by Heterogeneous Users". In: *Conf. on Learning Theory (COLT)*. 798–818.

Full text available at: http://dx.doi.org/10.1561/2200000068

Chen, T. and G. B. Giannakis. 2018. "Bandit convex optimization for scalable and dynamic IoT management". *IEEE Internet of Things Journal*.

Chen, T., Q. Ling, and G. B. Giannakis. 2017. "An online convex optimization approach to proactive network resource allocation". *IEEE Transactions on Signal Processing*. 65(24): 6350–6364.

Chen, W., Y. Wang, and Y. Yuan. 2013. "Combinatorial Multi-Armed Bandit: General Framework and Applications". In: *20th Intl. Conf. on Machine Learning (ICML)*. 151–159.

Chen, Y., C. Lee, H. Luo, and C. Wei. 2019. "A New Algorithm for Non-stationary Contextual Bandits: Efficient, Optimal, and Parameter-free". In: *Conf. on Learning Theory (COLT)*.

Cheung, Y. K. and G. Piliouras. 2019. "Vortices Instead of Equilibria in MinMax Optimization: Chaos and Butterfly Effects of Online Learning in Zero-Sum Games". In: *Conf. on Learning Theory (COLT)*. 807–834.

Chow, S.-C. and M. Chang. 2008. "Adaptive design methods in clinical trials – a review". *Orphanet Journal of Rare Diseases*. 3(11): 1750–1172.

Christiano, P., J. A. Kelner, A. Madry, D. A. Spielman, and S.-H. Teng. 2011. "Electrical Flows, Laplacian Systems, and Faster Approximation of Maximum Flow in Undirected Graphs". In: *43rd ACM Symp. on Theory of Computing (STOC)*. ACM. 273–282.

Chu, W., L. Li, L. Reyzin, and R. E. Schapire. 2011. "Contextual Bandits with Linear Payoff Functions". In: *14th Intl. Conf. on Artificial Intelligence and Statistics (AISTATS)*.

Cohen, L. and Y. Mansour. 2019. "Optimal Algorithm for Bayesian Incentive-Compatible Exploration". In: *ACM Conf. on Economics and Computation (ACM EC)*. 135–151.

Combes, R., C. Jiang, and R. Srikant. 2015. "Bandits with budgets: Regret lower bounds and optimal algorithms". *ACM SIGMETRICS Performance Evaluation Review*. 43(1): 245–257.

Cover, T. M. and J. A. Thomas. 1991. *Elements of Information Theory*. New York: John Wiley & Sons.

Full text available at: http://dx.doi.org/10.1561/2200000068

Dani, V., T. P. Hayes, and S. Kakade. 2008. "Stochastic Linear Optimization under Bandit Feedback". In: *21th Conf. on Learning Theory (COLT)*. 355–366.

Daskalakis, C., A. Deckelbaum, and A. Kim. 2015. "Near-optimal no-regret algorithms for zero-sum games". *Games and Economic Behavior*. 92: 327–348. Preliminary version in *ACM-SIAM SODA 2011*.

Daskalakis, C., A. Ilyas, V. Syrgkanis, and H. Zeng. 2018. "Training GANs with Optimism". In: *6th International Conference on Learning Representations (ICLR)*.

Daskalakis, C. and Q. Pan. 2014. "A Counter-example to Karlin's Strong Conjecture for Fictitious Play". In: *55th IEEE Symp. on Foundations of Computer Science (FOCS)*. 11–20.

Dekel, O., A. Tewari, and R. Arora. 2012. "Online Bandit Learning against an Adaptive Adversary: from Regret to Policy Regret". In: *29th Intl. Conf. on Machine Learning (ICML)*.

Desautels, T., A. Krause, and J. Burdick. 2012. "Parallelizing Exploration-Exploitation Tradeoffs with Gaussian Process Bandit Optimization". In: *29th Intl. Conf. on Machine Learning (ICML)*.

Devanur, N. R. and T. P. Hayes. 2009. "The AdWords problem: Online keyword matching with budgeted bidders under random permutations". In: *10th ACM Conf. on Electronic Commerce (EC)*. 71–78.

Devanur, N. R., K. Jain, B. Sivan, and C. A. Wilkens. 2011. "Near optimal online algorithms and fast approximation algorithms for resource allocation problems". In: *12th ACM Conf. on Electronic Commerce (EC)*. 29–38.

Devanur, N. R., K. Jain, B. Sivan, and C. A. Wilkens. 2019. "Near Optimal Online Algorithms and Fast Approximation Algorithms for Resource Allocation Problems". *J. ACM*. 66(1): 7:1–7:41. Preliminary version in *ACM EC 2011*.

Devanur, N. and S. M. Kakade. 2009. "The Price of Truthfulness for Pay-Per-Click Auctions". In: *10th ACM Conf. on Electronic Commerce (EC)*. 99–106.

Ding, W., T. Qin, X.-D. Zhang, and T.-Y. Liu. 2013. "Multi-Armed Bandit with Budget Constraint and Variable Costs". In: *27th AAAI Conference on Artificial Intelligence (AAAI)*.

Full text available at: http://dx.doi.org/10.1561/2200000068

Dong, M., Q. Li, D. Zarchy, P. B. Godfrey, and M. Schapira. 2015. "PCC: Re-architecting Congestion Control for Consistent High Performance". In: *12th USENIX Symp. on Networked Systems Design and Implementation (NSDI)*. 395–408.

Dong, M., T. Meng, D. Zarchy, E. Arslan, Y. Gilad, B. Godfrey, and M. Schapira. 2018. "PCC Vivace: Online-Learning Congestion Control". In: *15th USENIX Symp. on Networked Systems Design and Implementation (NSDI)*. 343–356.

Dubhashi, D. P. and A. Panconesi. 2009. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press.

Dudik, M., D. Erhan, J. Langford, and L. Li. 2012. "Sample-efficient Nonstationary Policy Evaluation for Contextual Bandits". In: *28th Conf. on Uncertainty in Artificial Intelligence (UAI)*. 247–254.

Dudik, M., D. Erhan, J. Langford, and L. Li. 2014. "Doubly Robust Policy Evaluation and Optimization". *Statistical Science*. 29(4): 1097–1104.

Dudik, M., N. Haghtalab, H. Luo, R. E. Schapire, V. Syrgkanis, and J. W. Vaughan. 2017. "Oracle-Efficient Online Learning and Auction Design". In: *58th IEEE Symp. on Foundations of Computer Science (FOCS)*. 528–539.

Dudik, M., K. Hofmann, R. E. Schapire, A. Slivkins, and M. Zoghi. 2015. "Contextual Dueling Bandits". In: *28th Conf. on Learning Theory (COLT)*.

Dudik, M., D. Hsu, S. Kale, N. Karampatziakis, J. Langford, L. Reyzin, and T. Zhang. 2011. "Efficient Optimal Leanring for Contextual Bandits". In: *27th Conf. on Uncertainty in Artificial Intelligence (UAI)*.

Even-Dar, E., S. Mannor, and Y. Mansour. 2002. "PAC bounds for multi-armed bandit and Markov decision processes". In: *15th Conf. on Learning Theory (COLT)*. 255–270.

Even-Dar, E., S. Mannor, and Y. Mansour. 2006. "Action Elimination and Stopping Conditions for the Multi-Armed Bandit and Reinforcement Learning Problems". *J. of Machine Learning Research (JMLR)*. 7: 1079–1105.

Full text available at: http://dx.doi.org/10.1561/2200000068

Feige, U., T. Koren, and M. Tennenholtz. 2017. "Chasing Ghosts: Competing with Stateful Policies". *SIAM J. on Computing (SICOMP)*. 46(1): 190–223. Preliminary version in *IEEE FOCS 2014*.

Feldman, J., M. Henzinger, N. Korula, V. S. Mirrokni, and C. Stein. 2010. "Online Stochastic Packing Applied to Display Ad Allocation". In: *18th Annual European Symp. on Algorithms (ESA)*. 182–194.

Flaxman, A., A. Kalai, and H. B. McMahan. 2005. "Online Convex Optimization in the Bandit Setting: Gradient Descent without a Gradient". In: *16th ACM-SIAM Symp. on Discrete Algorithms (SODA)*. 385–394.

Foster, D. and R. Vohra. 1997. "Calibrated learning and correlated equilibrium". *Games and Economic Behavior*. 21: 40–55.

Foster, D. and R. Vohra. 1998. "Asymptotic calibration". *Biometrika*. 85: 379–390.

Foster, D. and R. Vohra. 1999. "Regret in the on-line decision problem". *Games and Economic Behavior*. 29: 7–36.

Foster, D. J., A. Agarwal, M. Dudik, H. Luo, and R. E. Schapire. 2018. "Practical Contextual Bandits with Regression Oracles". In: *35th Intl. Conf. on Machine Learning (ICML)*. 1534–1543.

Foster, D. J., Z. Li, T. Lykouris, K. Sridharan, and É. Tardos. 2016. "Learning in Games: Robustness of Fast Convergence". In: *29th Advances in Neural Information Processing Systems (NIPS)*. 4727–4735.

Frazier, P., D. Kempe, J. M. Kleinberg, and R. Kleinberg. 2014. "Incentivizing exploration". In: *ACM Conf. on Economics and Computation (ACM EC)*. 5–22.

Freund, Y. and R. E. E. Schapire. 1997. "A decision-theoretic generalization of on-line learning and an application to boosting". *Journal of Computer and System Sciences*. 55(1): 119–139.

Freund, Y. and R. E. Schapire. 1996. "Game theory, on-line prediction and boosting". In: *9th Conf. on Learning Theory (COLT)*. 325–332.

Freund, Y. and R. E. Schapire. 1999. "Adaptive game playing using multiplicative weights". *Games and Economic Behavior*. 29(1-2): 79–103.

Full text available at: http://dx.doi.org/10.1561/2200000068

Freund, Y., R. E. Schapire, Y. Singer, and M. K. Warmuth. 1997. "Using and combining predictors that specialize". In: *29th ACM Symp. on Theory of Computing (STOC)*. 334–343.

Garivier, A. and O. Cappé. 2011. "The KL-UCB Algorithm for Bounded Stochastic Bandits and Beyond". In: *24th Conf. on Learning Theory (COLT)*.

Garivier, A. and E. Moulines. 2011. "On Upper-Confidence Bound Policies for Switching Bandit Problems". In: *22nd Intl. Conf. on Algorithmic Learning Theory (ALT)*. 174–188.

Gatti, N., A. Lazaric, and F. Trovo. 2012. "A Truthful Learning Mechanism for Contextual Multi-Slot Sponsored Search Auctions with Externalities". In: *13th ACM Conf. on Electronic Commerce (EC)*.

Ghosh, A. and P. Hummel. 2013. "Learning and incentives in user-generated content: multi-armed bandits with endogenous arms". In: *Innovations in Theoretical Computer Science Conf. (ITCS)*. 233–246.

Gittins, J. C. 1979. "Bandit processes and dynamic allocation indices (with discussion)". *J. Roy. Statist. Soc. Ser. B.* 41: 148–177.

Gittins, J., K. Glazebrook, and R. Weber. 2011. *Multi-Armed Bandit Allocation Indices*. John Wiley & Sons.

Golovin, D., A. Krause, and M. Streeter. 2009. "Online Learning of Assignments". In: *Advances in Neural Information Processing Systems (NIPS)*.

Grill, J., M. Valko, and R. Munos. 2015. "Black-box optimization of noisy functions with unknown smoothness". In: *28th Advances in Neural Information Processing Systems (NIPS)*.

Guha, S. and K. Munagala. 2007. "Multi-armed Bandits with Metric Switching Costs". In: *36th Intl. Colloquium on Automata, Languages and Programming (ICALP)*. 496–507.

Gupta, A., T. Koren, and K. Talwar. 2019. "Better Algorithms for Stochastic Bandits with Adversarial Corruptions". In: *Conf. on Learning Theory (COLT)*. 1562–1578.

Gupta, A., R. Krauthgamer, and J. R. Lee. 2003. "Bounded Geometries, Fractals, and Low–distortion Embeddings". In: *44th IEEE Symp. on Foundations of Computer Science (FOCS)*. 534–543.

Full text available at: http://dx.doi.org/10.1561/2200000068

Gupta, A., R. Krishnaswamy, M. Molinaro, and R. Ravi. 2011. "Approximation Algorithms for Correlated Knapsacks and Non-martingale Bandits". In: *52nd IEEE Symp. on Foundations of Computer Science (FOCS)*. 827–836.

György, A., L. Kocsis, I. Szabó, and C. Szepesvári. 2007a. "Continuous Time Associative Bandit Problems". In: *20th Intl. Joint Conf. on Artificial Intelligence (IJCAI)*. 830–835.

György, A., T. Linder, G. Lugosi, and G. Ottucsák. 2007b. "The On-Line Shortest Path Problem Under Partial Monitoring". *J. of Machine Learning Research (JMLR)*. 8: 2369–2403.

Hannan, J. 1957. "Approximation to Bayes risk in repeated play". *Contributions to the Theory of Games*. 3: 97–139.

Hart, S. and A. Mas-Colell. 2000. "A simple adaptive procedure leading to correlated equilibrium". *Econometrica*. 68: 1127–1150.

Hazan, E. 2015. "Introduction to Online Convex Optimization". *Foundations and Trends® in Optimization*. 2(3-4): 157–325.

Hazan, E. and S. Kale. 2011. "Better algorithms for benign bandits". *Journal of Machine Learning Research*. 12: 1287–1311. Preliminary version published in *ACM-SIAM SODA 2009*.

Hazan, E. and K. Y. Levy. 2014. "Bandit Convex Optimization: Towards Tight Bounds". In: *27th Advances in Neural Information Processing Systems (NIPS)*. 784–792.

Hazan, E. and N. Megiddo. 2007. "Online Learning with Prior Information". In: *20th Conf. on Learning Theory (COLT)*. 499–513.

Heidari, H., M. Mahdian, U. Syed, S. Vassilvitskii, and S. Yazdanbod. 2016. "Pricing a Low-regret Seller". In: *33rd Intl. Conf. on Machine Learning (ICML)*. 2559–2567.

Ho, C.-J., A. Slivkins, and J. W. Vaughan. 2016. "Adaptive Contract Design for Crowdsourcing Markets: Bandit Algorithms for Repeated Principal-Agent Problems". *J. of Artificial Intelligence Research*. 55: 317–359. Preliminary version appeared in *ACM EC 2014*.

Hofmann, K., L. Li, and F. Radlinski. 2016. "Online Evaluation for Information Retrieval". *Foundations and Trends® in Information Retrieval*. 10(1): 1–117.

Full text available at: http://dx.doi.org/10.1561/2200000068

Honda, J. and A. Takemura. 2010. "An Asymptotically Optimal Bandit Algorithm for Bounded Support Models". In: *23rd Conf. on Learning Theory (COLT)*.

Hsu, J., Z. Huang, A. Roth, and Z. S. Wu. 2016. "Jointly private convex programming". In: *27th ACM-SIAM Symp. on Discrete Algorithms (SODA)*. 580–599.

Immorlica, N., J. Mao, A. Slivkins, and S. Wu. 2018. "Incentivizing Exploration with Unbiased History". Working paper. URL: https://arxiv.org/abs/1811.06026.

Immorlica, N., J. Mao, A. Slivkins, and S. Wu. 2019a. "Bayesian Exploration with Heterogenous Agents". In: *The Web Conference (formerly known as* WWW*)*.

Immorlica, N., K. A. Sankararaman, R. Schapire, and A. Slivkins. 2019b. "Adversarial Bandits with Knapsacks". In: *60th IEEE Symp. on Foundations of Computer Science (FOCS)*.

Jiang, J., R. Das, G. Ananthanarayanan, P. A. Chou, V. N. Padmanabhan, V. Sekar, E. Dominique, M. Goliszewski, D. Kukoleca, R. Vafin, and H. Zhang. 2016. "Via: Improving Internet Telephony Call Quality Using Predictive Relay Selection". In: *ACM SIGCOMM (ACM SIGCOMM Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications)*. 286–299.

Jiang, J., S. Sun, V. Sekar, and H. Zhang. 2017. "Pytheas: Enabling Data-Driven Quality of Experience Optimization Using Group-Based Exploration-Exploitation". In: *14th USENIX Symp. on Networked Systems Design and Implementation (NSDI)*. 393–406.

Kakade, S. M., I. Lobel, and H. Nazerzadeh. 2011. "Optimal Dynamic Mechanism Design and the Virtual Pivot Mechanism". SSRN Report, SSRN ID 1782211.

Kalai, A. T. and S. Vempala. 2005. "Efficient algorithms for online decision problems". *J. of Computer and Systems Sciences*. 71(3): 291–307. Preliminary version in *COLT 2003*.

Kale, S., L. Reyzin, and R. E. Schapire. 2010. "Non-Stochastic Bandit Slate Problems". In: *24th Advances in Neural Information Processing Systems (NIPS)*. 1054–1062.

Kamenica, E. and M. Gentzkow. 2011. "Bayesian Persuasion". *American Economic Review*. 101(6): 2590–2615.

Full text available at: http://dx.doi.org/10.1561/2200000068

Kannan, S., J. Morgenstern, A. Roth, B. Waggoner, and Z. S. Wu. 2018. "A Smoothed Analysis of the Greedy Algorithm for the Linear Contextual Bandit Problem". In: *Advances in Neural Information Processing Systems (NIPS)*.

Karger, D. and M. Ruhl. 2002. "Finding Nearest Neighbors in Growth-restricted Metrics". In: *34th ACM Symp. on Theory of Computing (STOC)*. 63–66.

Kaufmann, E., O. Cappé, and A. Garivier. 2016. "On the Complexity of Best-Arm Identification in Multi-Armed Bandit Models". *J. of Machine Learning Research (JMLR)*. 17: 1:1–1:42.

Kaufmann, E., N. Korda, and R. Munos. 2012. "Thompson Sampling: An Asymptotically Optimal Finite-Time Analysis". In: *23rd Intl. Conf. on Algorithmic Learning Theory (ALT)*. 199–213.

Kearns, M., S. Neel, A. Roth, and Z. S. Wu. 2018. "Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness". In: *35th Intl. Conf. on Machine Learning (ICML)*. 2564–2572.

Keller, G., S. Rady, and M. Cripps. 2005. "Strategic Experimentation with Exponential Bandits". *Econometrica*. 73(1): 39–68.

Kleinberg, J., A. Slivkins, and T. Wexler. 2009a. "Triangulation and Embedding Using Small Sets of Beacons". *J. of the ACM*. 56(6). Subsumes conference papers in *IEEE FOCS 2004* and *ACM-SIAM SODA 2005*.

Kleinberg, J. and E. Tardos. 2005. *Algorithm Design*. Addison Wesley.

Kleinberg, R. 2004. "Nearly Tight Bounds for the Continuum-Armed Bandit Problem." In: *18th Advances in Neural Information Processing Systems (NIPS)*.

Kleinberg, R. 2006. "Anytime algorithms for multi-armed bandit problems." In: *17th ACM-SIAM Symp. on Discrete Algorithms (SODA)*. 928–936.

Kleinberg, R. 2007. "*CS683: Learning, Games, and Electronic Markets*, a class at Cornell University". Lecture notes, available at http://www.cs.cornell.edu/courses/cs683/2007sp/.

Kleinberg, R. D. and F. T. Leighton. 2003. "The Value of Knowing a Demand Curve: Bounds on Regret for Online Posted-Price Auctions". In: *IEEE Symp. on Foundations of Computer Science (FOCS)*.

Full text available at: http://dx.doi.org/10.1561/2200000068

Kleinberg, R., A. Niculescu-Mizil, and Y. Sharma. 2008a. "Regret bounds for sleeping experts and bandits". In: *21st Conf. on Learning Theory (COLT)*. 425–436.

Kleinberg, R., G. Piliouras, and É. Tardos. 2009b. "Multiplicative updates outperform generic no-regret learning in congestion games: extended abstract". In: *41st ACM Symp. on Theory of Computing (STOC)*. 533–542.

Kleinberg, R. and A. Slivkins. 2010. "Sharp Dichotomies for Regret Minimization in Metric Spaces". In: *21st ACM-SIAM Symp. on Discrete Algorithms (SODA)*.

Kleinberg, R., A. Slivkins, and E. Upfal. 2008b. "Multi-Armed Bandits in Metric Spaces". In: *40th ACM Symp. on Theory of Computing (STOC)*. 681–690.

Kleinberg, R., A. Slivkins, and E. Upfal. 2019. "Bandits and Experts in Metric Spaces". *J. of the ACM*. 66(4). Merged and revised version of conference papers in *ACM STOC 2008* and *ACM-SIAM SODA 2010*. Also available at http://arxiv.org/abs/1312.1277.

Kocsis, L. and C. Szepesvari. 2006. "Bandit Based Monte-Carlo Planning". In: *17th European Conf. on Machine Learning (ECML)*. 282–293.

Koolen, W. M., M. K. Warmuth, and J. Kivinen. 2010. "Hedging Structured Concepts". In: *23rd Conf. on Learning Theory (COLT)*.

Krause, A. and C. S. Ong. 2011. "Contextual Gaussian Process Bandit Optimization". In: *25th Advances in Neural Information Processing Systems (NIPS)*. 2447–2455.

Kremer, I., Y. Mansour, and M. Perry. 2014. "Implementing the "Wisdom of the Crowd"". *J. of Political Economy*. 122(5): 988–1012. Preliminary version in *ACM EC 2014*.

Krishnamurthy, A., A. Agarwal, and M. Dudik. 2016. "Contextual semibandits via supervised learning oracles". In: *29th Advances in Neural Information Processing Systems (NIPS)*.

Krishnamurthy, A., J. Langford, A. Slivkins, and C. Zhang. 2019. "Contextual bandits with continuous actions: Smoothing, zooming, and adapting". In: *Conf. on Learning Theory (COLT)*. Working paper, under journal submission. URL: https://arxiv.org/abs/1902.01520.

Full text available at: http://dx.doi.org/10.1561/2200000068

Kveton, B., C. Szepesvári, Z. Wen, and A. Ashkan. 2015a. "Cascading Bandits: Learning to Rank in the Cascade Model". In: *32nd Intl. Conf. on Machine Learning (ICML)*. 767–776.

Kveton, B., Z. Wen, A. Ashkan, H. Eydgahi, and B. Eriksson. 2014a. "Matroid Bandits: Fast Combinatorial Optimization with Learning". In: *13th Conf. on Uncertainty in Artificial Intelligence (UAI)*. 420–429.

Kveton, B., Z. Wen, A. Ashkan, H. Eydgahi, and B. Eriksson. 2014b. "Matroid Bandits: Fast Combinatorial Optimization with Learning." In: *Conf. on Uncertainty in Artificial Intelligence (UAI)*. Ed. by N. L. Zhang and J. Tian. 420–429.

Kveton, B., Z. Wen, A. Ashkan, and C. Szepesvári. 2015b. "Combinatorial Cascading Bandits". In: *28th Advances in Neural Information Processing Systems (NIPS)*. 1450–1458.

Kveton, B., Z. Wen, A. Ashkan, and C. Szepesvári. 2015c. "Tight Regret Bounds for Stochastic Combinatorial Semi-Bandits". In: *18th Intl. Conf. on Artificial Intelligence and Statistics (AISTATS)*.

Laffont, J.-J. and D. Martimort. 2002. *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press.

Lai, L., H. Jiang, and H. V. Poor. 2008. "Medium access in cognitive radio networks: A competitive multi-armed bandit framework". In: *42nd Asilomar Conference on Signals, Systems and Computers*.

Lai, T. L. and H. Robbins. 1985. "Asymptotically efficient Adaptive Allocation Rules". *Advances in Applied Mathematics*. 6: 4–22.

Langford, J. and T. Zhang. 2007. "The Epoch-Greedy Algorithm for Contextual Multi-armed Bandits". In: *21st Advances in Neural Information Processing Systems (NIPS)*.

Lattimore, T. and C. Szepesvári. 2019. *Bandit Algorithms*. Cambridge University Press (preprint).

Li, L., S. Chen, J. Kleban, and A. Gupta. 2015. "Counterfactual Estimation and Optimization of Click Metrics in Search Engines: A Case Study". In: *24th Intl. World Wide Web Conf. (WWW)*. 929–934.

Li, L., W. Chu, J. Langford, and R. E. Schapire. 2010. "A contextual-bandit approach to personalized news article recommendation". In: *19th Intl. World Wide Web Conf. (WWW)*.

Full text available at: http://dx.doi.org/10.1561/2200000068

Li, L., W. Chu, J. Langford, and X. Wang. 2011. "Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms". In: *4th ACM Intl. Conf. on Web Search and Data Mining (WSDM)*.

Li, S., A. Karatzoglou, and C. Gentile. 2016. "Collaborative Filtering Bandits". In: *16th ACM Intl. Conf. on Research and Development in Information Retrieval (SIGIR)*. 539–548.

Littlestone, N. and M. K. Warmuth. 1994. "The Weighted Majority Algorithm". *Information and Computation*. 108(2): 212–260.

Liu, K. and Q. Zhao. 2010. "Distributed learning in multi-armed bandit with multiple players". *IEEE Trans. Signal Processing*. 58(11): 5667–5681.

Lu, T., D. Pál, and M. Pál. 2010. "Showing Relevant Ads via Lipschitz Context Multi-Armed Bandits". In: *14th Intl. Conf. on Artificial Intelligence and Statistics (AISTATS)*.

Luo, H., C. Wei, A. Agarwal, and J. Langford. 2018. "Efficient Contextual Bandits in Non-stationary Worlds". In: *Conf. on Learning Theory (COLT)*. 1739–1776.

Lykouris, T., V. Mirrokni, and R. Paes-Leme. 2018. "Stochastic bandits robust to adversarial corruptions". In: *50th ACM Symp. on Theory of Computing (STOC)*.

Lykouris, T., V. Syrgkanis, and É. Tardos. 2016. "Learning and Efficiency in Games with Dynamic Population". In: *27th ACM-SIAM Symp. on Discrete Algorithms (SODA)*. 120–129.

Magureanu, S., R. Combes, and A. Proutiere. 2014. "Lipschitz Bandits: Regret Lower Bound and Optimal Algorithms". In: *27th Conf. on Learning Theory (COLT)*. 975–999.

Mahdavi, M., R. Jin, and T. Yang. 2012. "Trading regret for efficiency: online convex optimization with long term constraints". *J. of Machine Learning Research (JMLR)*. 13(Sep): 2503–2528.

Mahdavi, M., T. Yang, and R. Jin. 2013. "Stochastic convex optimization with multiple objectives". In: *Advances in Neural Information Processing Systems (NIPS)*. 1115–1123.

Full text available at: http://dx.doi.org/10.1561/2200000068

Maillard, O.-A. and R. Munos. 2010. "Online Learning in Adversarial Lipschitz Environments". In: *European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*. 305–320.

Maillard, O., R. Munos, and G. Stoltz. 2011. "A Finite-Time Analysis of Multi-armed Bandits Problems with Kullback-Leibler Divergences". In: *24th Conf. on Learning Theory (COLT)*.

Mannor, S. and J. N. Tsitsiklis. 2004. "The sample complexity of exploration in the multi-armed bandit problem". *J. of Machine Learning Research (JMLR)*. 5: 623–648.

Mansour, Y., A. Slivkins, and V. Syrgkanis. 2019. "Bayesian Incentive-Compatible Bandit Exploration". *Operations Research*. To appear; preliminary version in *ACM EC 2015*.

Mansour, Y., A. Slivkins, V. Syrgkanis, and S. Wu. 2016. "Bayesian Exploration: Incentivizing Exploration in Bayesian Games". In: *16th ACM Conf. on Economics and Computation (ACM EC)*. To appear in *Operations Research* after a revision.

Mansour, Y., A. Slivkins, and S. Wu. 2018. "Competing Bandits: Learning under Competition". In: *9th Innovations in Theoretical Computer Science Conf. (ITCS)*.

McDiarmid, C. 1998. "Concentration". In: *Probabilistic Methods for Discrete Mathematics*. Ed. by M. H. C. M. J. Ramirez and B. Reed. Berlin: Springer-Verlag. 195–248.

McMahan, H. B. and A. Blum. 2004. "Online Geometric Optimization in the Bandit Setting Against an Adaptive Adversary". In: *17th Conf. on Learning Theory (COLT)*. 109–123.

Merhav, N., E. Ordentlich, G. Seroussi, and M. J. Weinberger. 2002. "On Sequential strategies for loss functions with memory". *IEEE Trans. on Information Theory*. 48(7): 1947–1958.

Mertikopoulos, P., C. H. Papadimitriou, and G. Piliouras. 2018. "Cycles in Adversarial Regularized Learning". In: *29th ACM-SIAM Symp. on Discrete Algorithms (SODA)*. 2703–2717.

Minsker, S. 2013. "Estimation of Extreme Values and Associated Level Sets of a Regression Function via Selective Sampling". In: *26th Conf. on Learning Theory (COLT)*. 105–121.

Full text available at: http://dx.doi.org/10.1561/2200000068

Molinaro, M. and R. Ravi. 2012. "Geometry of Online Packing Linear Programs". In: *39th Intl. Colloquium on Automata, Languages and Programming (ICALP)*. 701–713.

Moulin, H. and J.-P. Vial. 1978. "Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon". *Intl. J. of Game Theory*. 7(3): 201–221.

Munos, R. 2011. "Optimistic Optimization of a Deterministic Function without the Knowledge of its Smoothness". In: *25th Advances in Neural Information Processing Systems (NIPS)*. 783–791.

Munos, R. 2014. "From Bandits to Monte-Carlo Tree Search: The Optimistic Principle Applied to Optimization and Planning". *Foundations and Trends in Machine Learning*. 7(1): 1–129.

Munos, R. and P.-A. Coquelin. 2007. "Bandit algorithms for tree search". In: *23rd Conf. on Uncertainty in Artificial Intelligence (UAI)*.

Nazerzadeh, H., A. Saberi, and R. Vohra. 2013. "Dynamic Pay-Per-Action Mechanisms and Applications to Online Advertising". *Operations Research*. 61(1): 98–111. Preliminary version in *WWW 2008*.

Neely, M. J. and H. Yu. 2017. "Online convex optimization with time-varying constraints". *arXiv preprint*. arXiv: 1702.04783.

Nekipelov, D., V. Syrgkanis, and É. Tardos. 2015. "Econometrics for Learning Agents". In: *16th ACM Conf. on Electronic Commerce (EC)*. 1–18.

Nisan, N. and G. Noti. 2017. "An Experimental Evaluation of Regret-Based Econometrics". In: *26th Intl. World Wide Web Conf. (WWW)*. 73–81.

Pandey, S., D. Agarwal, D. Chakrabarti, and V. Josifovski. 2007a. "Bandits for Taxonomies: A Model-based Approach". In: *SIAM Intl. Conf. on Data Mining (SDM)*.

Pandey, S., D. Chakrabarti, and D. Agarwal. 2007b. "Multi-armed Bandit Problems with Dependent Arms". In: *24th Intl. Conf. on Machine Learning (ICML)*.

Pavan, A., I. Segal, and J. Toikka. 2011. "Dynamic Mechanism Design: Revenue Equivalence, Profit Maximization, and Information Disclosure". Working paper.

Full text available at: http://dx.doi.org/10.1561/2200000068

Radlinski, F., R. Kleinberg, and T. Joachims. 2008. "Learning diverse rankings with multi-armed bandits". In: *25th Intl. Conf. on Machine Learning (ICML)*. 784–791.

Raghavan, M., A. Slivkins, J. W. Vaughan, and Z. S. Wu. 2018. "The Externalities of Exploration and How Data Diversity Helps Exploitation". In: *Conf. on Learning Theory (COLT)*. 1724–1738.

Rakhlin, A. and K. Sridharan. 2013. "Optimization, Learning, and Games with Predictable Sequences". In: *27th Advances in Neural Information Processing Systems (NIPS)*. 3066–3074.

Rakhlin, A. and K. Sridharan. 2016. "BISTRO: An Efficient Relaxation-Based Method for Contextual Bandits". In: *33nd Intl. Conf. on Machine Learning (ICML)*.

Rakhlin, A., K. Sridharan, and A. Tewari. 2015. "Online learning via sequential complexities". *J. of Machine Learning Research (JMLR)*. 16: 155–186.

Rangi, A., M. Franceschetti, and L. Tran-Thanh. 2019. "Unifying the Stochastic and the Adversarial Bandits with Knapsack". In: *28th Intl. Joint Conf. on Artificial Intelligence (IJCAI)*. 3311–3317.

Rivera, A., H. Wang, and H. Xu. 2018. "Online Saddle Point Problem with Applications to Constrained Online Convex Optimization". *arXiv preprint*. arXiv: 1806.08301.

Robinson, J. 1951. "An iterative method of solving a game". *Annals of Mathematics, Second Series*. 54(2): 296–301.

Rogers, R., A. Roth, J. Ullman, and Z. S. Wu. 2015. "Inducing approximately optimal flow using truthful mediators". In: *16th ACM Conf. on Electronic Commerce (EC)*. 471–488.

Rosenski, J., O. Shamir, and L. Szlak. 2016. "Multi-Player Bandits - a Musical Chairs Approach". In: *33nd Intl. Conf. on Machine Learning (ICML)*. 155–163.

Roth, A., A. Slivkins, J. Ullman, and Z. S. Wu. 2017. "Multidimensional dynamic pricing for welfare maximization". In: *18th ACM Conf. on Electronic Commerce (EC)*. 519–536.

Roth, A., J. Ullman, and Z. S. Wu. 2016. "Watch and learn: Optimizing from revealed preferences feedback". In: *48th ACM Symp. on Theory of Computing (STOC)*. 949–962.

Full text available at: http://dx.doi.org/10.1561/2200000068

Roughgarden, T. 2009. "Intrinsic robustness of the price of anarchy". In: *41st ACM Symp. on Theory of Computing (STOC)*. 513–522.

Roughgarden, T. 2016. *Twenty Lectures on Algorithmic Game Theory*. Cambridge University Press.

Rusmevichientong, P. and J. N. Tsitsiklis. 2010. "Linearly Parameterized Bandits". *Mathematics of Operations Research*. 35(2): 395–411.

Russo, D. and B. V. Roy. 2014. "Learning to Optimize via Posterior Sampling". *Mathematics of Operations Research*. 39(4): 1221–1243.

Russo, D. and B. V. Roy. 2016. "An Information-Theoretic Analysis of Thompson Sampling". *J. of Machine Learning Research (JMLR)*. 17: 68:1–68:30.

Russo, D., B. V. Roy, A. Kazerouni, I. Osband, and Z. Wen. 2018. "A Tutorial on Thompson Sampling". *Foundations and Trends in Machine Learning*. 11(1): 1–96.

Sankararaman, K. A. and A. Slivkins. 2018. "Combinatorial Semi-Bandits with Knapsacks". In: *Intl. Conf. on Artificial Intelligence and Statistics (AISTATS)*. 1760–1770.

Sauré, D. and A. Zeevi. 2013. "Optimal Dynamic Assortment Planning with Demand Learning". *Manufacturing & Service Operations Management*. 15(3): 387–404.

Schmit, S. and C. Riquelme. 2018. "Human Interaction with Recommendation Systems". In: *Intl. Conf. on Artificial Intelligence and Statistics (AISTATS)*. 862–870.

Schroeder, M. 1991. *Fractal, Chaos and Power Laws: Minutes from an Infinite Paradise*. W. H. Freeman and Co.

Schumpeter, J. 1942. *Capitalism, Socialism and Democracy*. Harper & Brothers.

Seldin, Y. and G. Lugosi. 2016. "A lower bound for multi-armed bandits with expert advice". In: *13th European Workshop on Reinforcement Learning (EWRL)*.

Seldin, Y. and G. Lugosi. 2017. "An Improved Parametrization and Analysis of the EXP3++ Algorithm for Stochastic and Adversarial Bandits". In: *30th Conf. on Learning Theory (COLT)*.

Seldin, Y. and A. Slivkins. 2014. "One Practical Algorithm for Both Stochastic and Adversarial Bandits". In: *31th Intl. Conf. on Machine Learning (ICML)*.

Full text available at: http://dx.doi.org/10.1561/2200000068

Sellke, M. 2019. Personal communication.

Sellke, M. and A. Slivkins. 2019. "Advances in Incentivized Exploration [tentative title]". Working paper.

Shamir, O. 2015. "On the Complexity of Bandit Linear Optimization". In: *28th Conf. on Learning Theory (COLT)*. 1523–1551.

Singla, A. and A. Krause. 2013. "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms". In: *22nd Intl. World Wide Web Conf. (WWW)*. 1167–1178.

Slivkins, A. 2007. "Towards Fast Decentralized Construction of Locality-Aware Overlay Networks". In: *26th Annual ACM Symp. on Principles Of Distributed Computing (PODC)*. 89–98.

Slivkins, A. 2011. "Multi-armed bandits on implicit metric spaces". In: *25th Advances in Neural Information Processing Systems (NIPS)*.

Slivkins, A. 2013. "Dynamic Ad Allocation: Bandits with Budgets". A technical report on arxiv.org/abs/1306.0155.

Slivkins, A. 2014. "Contextual bandits with similarity information". *J. of Machine Learning Research (JMLR)*. 15(1): 2533–2568. Preliminary version in *COLT 2011*.

Slivkins, A. 2017. "Incentivizing exploration via information asymmetry". *ACM Crossroads*. 24(1): 38–41.

Slivkins, A. and E. Upfal. 2008. "Adapting to a Changing Environment: the Brownian Restless Bandits". In: *21st Conf. on Learning Theory (COLT)*. 343–354.

Slivkins, A. and J. W. Vaughan. 2013. "Online Decision Making in Crowdsourcing Markets: Theoretical Challenges". *SIGecom Exchanges*. 12(2).

Srinivas, N., A. Krause, S. Kakade, and M. Seeger. 2010. "Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design". In: *27th Intl. Conf. on Machine Learning (ICML)*. 1015–1022.

Stoltz, G. 2005. "Incomplete Information and Internal Regret in Prediction of Individual Sequences". *PhD thesis*. University Paris XI, ORSAY.

Streeter, M. and D. Golovin. 2008. "An Online Algorithm for Maximizing Submodular Functions". In: *Advances in Neural Information Processing Systems (NIPS)*. 1577–1584.

Full text available at: http://dx.doi.org/10.1561/2200000068

Sutton, R. S. and A. G. Barto. 1998. *Reinforcement Learning: An Introduction*. MIT Press.

Swaminathan, A. and T. Joachims. 2015. "Batch learning from logged bandit feedback through counterfactual risk minimization". *J. of Machine Learning Research (JMLR)*. 16: 1731–1755.

Swaminathan, A., A. Krishnamurthy, A. Agarwal, M. Dudik, J. Langford, D. Jose, and I. Zitouni. 2017. "Off-policy evaluation for slate recommendation". In: *30th Advances in Neural Information Processing Systems (NIPS)*. 3635–3645.

Syrgkanis, V., A. Agarwal, H. Luo, and R. E. Schapire. 2015. "Fast Convergence of Regularized Learning in Games". In: *28th Advances in Neural Information Processing Systems (NIPS)*. 2989–2997.

Syrgkanis, V., A. Krishnamurthy, and R. E. Schapire. 2016a. "Efficient Algorithms for Adversarial Contextual Learning". In: *33nd Intl. Conf. on Machine Learning (ICML)*.

Syrgkanis, V., H. Luo, A. Krishnamurthy, and R. E. Schapire. 2016b. "Improved Regret Bounds for Oracle-Based Adversarial Contextual Bandits". In: *29th Advances in Neural Information Processing Systems (NIPS)*.

Syrgkanis, V. and É. Tardos. 2013. "Composable and efficient mechanisms". In: *45th ACM Symp. on Theory of Computing (STOC)*. 211–220.

Szepesvári, C. 2010. *Algorithms for Reinforcement Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan & Claypool Publishers.

Talwar, K. 2004. "Bypassing the embedding: Algorithms for Low-Dimensional Metrics". In: *36th ACM Symp. on Theory of Computing (STOC)*. 281–290.

Thompson, W. R. 1933. "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples." *Biometrika*. 25(3-4): 285–294.

Tran-Thanh, L., A. Chapman, E. M. de Cote, A. Rogers, and N. R. R. Jennings. 2010. "ϵ-first policies for budget-limited multi-armed bandits". In: *24th AAAI Conference on Artificial Intelligence (AAAI)*. 1211–1216.

Full text available at: http://dx.doi.org/10.1561/2200000068

Tran-Thanh, L., A. Chapman, A. Rogers, and N. R. R. Jennings. 2012. "Knapsack based optimal policies for budget-limited multi-armed bandits". In: *26th AAAI Conference on Artificial Intelligence (AAAI)*. 1134–1140.

Valko, M., A. Carpentier, and R. Munos. 2013. "Stochastic Simultaneous Optimistic Optimization". In: *30th Intl. Conf. on Machine Learning (ICML)*. 19–27.

Wang, J. and J. D. Abernethy. 2018. "Acceleration through Optimistic No-Regret Dynamics". In: *31st Advances in Neural Information Processing Systems (NIPS)*. 3828–3838.

Wang, Z., S. Deng, and Y. Ye. 2014. "Close the Gaps: A Learning-While-Doing Algorithm for Single-Product Revenue Management Problems". *Operations Research*. 62(2): 318–331.

Wei, C. and H. Luo. 2018. "More Adaptive Algorithms for Adversarial Bandits". In: *31st Conf. on Learning Theory (COLT)*.

Wilkens, C. and B. Sivan. 2012. "Single-Call Mechanisms". In: *13th ACM Conf. on Electronic Commerce (EC)*.

Yue, Y., J. Broder, R. Kleinberg, and T. Joachims. 2012. "The K-armed dueling bandits problem". *J. Comput. Syst. Sci.* 78(5): 1538–1556. Preliminary version in COLT 2009.

Yue, Y. and T. Joachims. 2009. "Interactively optimizing information retrieval systems as a dueling bandits problem". In: *26th Intl. Conf. on Machine Learning (ICML)*. 1201–1208.

Zimmert, J. and T. Lattimore. 2019. "Connections Between Mirror Descent, Thompson Sampling and the Information Ratio". In: *33rd Advances in Neural Information Processing Systems (NeurIPS)*.

Zimmert, J., H. Luo, and C. Wei. 2019. "Beating Stochastic and Adversarial Semi-bandits Optimally and Simultaneously". In: *36th Intl. Conf. on Machine Learning (ICML)*. 7683–7692.

Zimmert, J. and Y. Seldin. 2019. "An Optimal Algorithm for Stochastic and Adversarial Bandits". In: *Intl. Conf. on Artificial Intelligence and Statistics (AISTATS)*.

Zoghi, M., S. Whiteson, R. Munos, and M. de Rijke. 2014. "Relative Upper Confidence Bound for the K-Armed Dueling Bandits Problem". In: *Intl. Conf. on Machine Learning (ICML)*. 10–18.

Full text available at: http://dx.doi.org/10.1561/2200000068

Zong, S., H. Ni, K. Sung, N. R. Ke, Z. Wen, and B. Kveton. 2016. "Cascading Bandits for Large-Scale Recommendation Problems". In: *32nd Conf. on Uncertainty in Artificial Intelligence (UAI)*.