Corporate Social Responsibility via Multi-Armed Bandits

Tom Ron[*1], **Omer Ben-Porat**[*1], **Uri Shalit**[1]
[1]Technion - Israel Institute of Technology
{ront, omerbp}@campus.technion.ac.il, urishalit@technion.ac.il

Abstract

We propose a multi-armed bandit setting where each arm corresponds to a subpopulation, and pulling an arm is equivalent to granting an opportunity to this subpopulation. In this setting the decision-maker's fairness policy governs the number of opportunities each subpopulation should receive, which typically depends on the (unknown) reward from granting an opportunity to this subpopulation. The decision-maker can decide whether to provide these opportunities or pay a predefined monetary value for every withheld opportunity. The decision-maker's objective is to maximize her utility, which is the sum of rewards minus the cost of withheld opportunities. We provide a no-regret algorithm that maximizes the decision-maker's utility and complement our analysis with an almost-tight lower bound. Full version of the paper is available at https://tinyurl.com/y7s9avud.

1 Introduction

Algorithmic decision making plays a fundamental role in many facets of our lives; criminal justice [Berk, 2012, Berk et al., 2019, Northpointe, 2015], banking [Usi, Fuster et al., 2018, Pérez-Martín et al., 2018, Zhang et al., 2015], online-advertisement [McMahan et al., 2013, Oentaryo et al., 2014], hiring [Ama, How, Abel, 2015, Ajunwa and Greene, 2019] , and college admission [Acharya and Sinha, 2014, Lux et al., 2016, Waters and Miikkulainen, 2014] are just a few examples. With the abundance of applications in which algorithms operate, concerns about their ethics, fairness, and privacy have emerged. For instance, classification algorithms that were deemed to be unfair and discriminate based on factors like gender, race, and more [Dwork et al., 2012, Hardt et al., 2016, Zafar et al., 2015, Zhao et al., 2017]. Algorithmic fairness is a framework that, among other means, is aimed at ensuring the long-term welfare of such subpopulations when subject to algorithmic decision making.

Consider the following online advertisement use-case. A company wants to publish a job ad online and optimizes its campaign based on the cost-per-click. As witnessed by Lambrecht and Tucker [2018], women are less likely to see job ads for STEM positions since they have higher cost-per-click than men. If women are not exposed to information about STEM career opportunities, they may never apply to such jobs [Diekman et al., 2010]. In order to act fairly and display ads to all the subpopulations the company will need to sacrifice part of its short-term utility and pay a higher cost-per-click. Our goal in this paper is to better understand the trade-off that companies who wish to ensure their algorithms are more equitable face.

We focus on exploring the cost of fairness versus the cost of alternatives such as Corporate Social Responsibility [Carroll et al., 1991] (CSR hereinafter). CSR is an approach towards the goal of long-term welfare which is becoming increasingly popular among tech-giants these days. CSR is a self-regulation act of philanthropic responsibility in response to the rising concerns on ethical issues in businesses. For example, in 2019, Microsoft spent more than three billion dollars with minority, disabled, veteran, LGBTQ, and woman-owned businesses[1].

In this paper, we suggest an algorithmic approach to CSR in the setting of sequential decision making. Sequential decision making is often modeled as Multi-armed bandit problems (hereinafter MAB; see Auer et al. [2002] for a brief introduction). MABs enjoy massive commercial success and have myriad real-world applications [Chow and Chang, 2008, Fu, 2016, White, 2012, Zeng et al., 2016]. It is therefore unsurprising that fair aspects of MAB are examined. In this work we treat arms as subpopulations, and require that subpopulations would not starve from lack of *opportunities*. Opportunities can be granted to a subpopulations explicitly, i.e., by pulling the subpopulation's arm, or implicitly via CSR channels. Given the example above, companies have the choice whether to display ads to subpopulations with higher cost-per-click or to invest money in organizations that promote the long term well-being of those subpopulations.

We highlight the tension between the decision-maker that wants to maximize her reward and the cost of CSR. We consider the bandit reward to be the benefit derived from granting the opportunity to the subpopulation represented by the arm. For simplicity we use the term expected reward from here

[*]Equal contribution

[1]https://aka.ms/2019CSRReport

on. The amount of opportunities depends on how fairness is perceived by the decision-maker and the expected rewards. Unfortunately, information about the expected rewards is not known in advance and has to be explored by the decision-maker. We take a utilitarian approach: The utility of the decision-maker is composed of the rewards, clicks on displayed ads, and a transfer cost. The transfer cost is the amount the decision-maker invests in CSR for every deferred opportunity. Knowing the transfer cost in advance, the decision-maker can make an informed decision on how to allocate its resources. Our model casts light on the trade-off between the cost of opportunity and the cost of transferring the opportunity requirement to an external source.

1.1 Our Contribution

Our contribution is two-fold: technical and conceptual. Technically, we consider the typical MAB setting with K Bernoulli arms with horizon T and expectation vector $\boldsymbol{\mu}$, which is unknown. In addition, we introduce a *fairness function* f, $f : [0,1]^K \to [0,1]^K$, which determines the minimal number of pulls for each arm given the expected reward vector $\boldsymbol{\mu}$. The term $T \cdot f(\boldsymbol{\mu})_i$ quantifies the amount of *opportunities* subpopulation i deserves, which is a function of its own expected reward and the expected rewards of the other subpopulations. The decision-maker gains rewards, but pays a transfer cost of λ for every round of unmet opportunity. We assume that both f and λ are known in advance. We characterize the optimal algorithm that achieves a sub-linear regret of $\tilde{O}(T^{2/3})$, and show a matching lower bound. In the appendix, we augment our theoretical analysis with experimental, examining the implications of different fairness functions f and values of λ.

On the conceptual side, our framework reflects the trade-off between monetary rewards and subpopulation opportunities, which can be viewed as a means of providing long-term welfare. This perspective follows, e.g., self-regulation in revenue-driven commercial companies (as decision-makers) contributing to societal goals, or a policy maker that ensures that the decision-maker is fairness aware. In the former, sufficient opportunities are a CSR [Garriga and Melé, 2004] that is integrated in the company's objective by design. In the latter, the decision-maker provides opportunities explicitly by arm pulls, or implicitly by payments that are invested in that subpopulation by the policy maker (for, e.g., better computer labs in public schools). Crucially, the number of required opportunities depends on the expected rewards, known only in hindsight.

1.2 Related Work

Multi-armed bandit has been a fertile ground for many fairness-related application [Joseph et al., 2016a,b, Liu et al., 2017, Patil et al., 2019]. Joseph et al. [2016a,b] study fairness in MABs from the eyes of the decision-maker. We study fairness from the perspective of the arms and view arm pulling as granting an opportunity. This view was also adopted by Liu et al. [2017]. The work most related to ours is Patil et al. [2019]. The authors define fairness as pulling each arm at least a minimal number of times according to a predefined vector (where each entry corresponds to a subpopulation). The predefined vector is given by the policy maker and is independent of the subpopulation properties. However, our work differs from Patil et al. [2019] in two crucial aspects. First, while Patil et al. [2019] model fairness as a hard constraint, we better address real-world applications and treat it as a soft one. Our utilitarian approach, which is well-studied in economic contexts [Mas-Colell et al., 1995, Varian and Varian, 1992], accounts for trading rewards with opportunities. If providing opportunities explicitly by pulling the arms is financially unbearable, the decision-maker can do that implicitly by monetary transfers. Second, Patil et al. [2019] construct the fairness constraint by a predefined vector, while in our work the opportunity requirements depend on the expected rewards, which is only known in hindsight. This uncertainty exacerbate the problem even further . These differences and others lead to a lower bound of $\tilde{O}(T^{2/3})$ compared to a $\tilde{O}(\sqrt{T})$ in theirs.

2 Model

We consider a stochastic bandit problem; a decision-maker is given K arms, and pulls one at each time step $t = 1, 2, \ldots, T$. We denote by i_t the arm pulled at time t. When arm i is pulled at time t, the decision-maker receives a random reward, $r_t \sim \mathcal{D}_i$. We assume that for every $i \in [K]$, the reward distribution $\mathcal{D}_i$ is a Bernoulli distribution with expected value μ_i. This is without loss of generality, since we can reduce any instance with general $[0,1]$-supported distribution to an instance with Bernoulli arms using the technique of Agrawal and Goyal [2012]. We use $\boldsymbol{\mu}$ to denote the vector of expected rewards, i.e., $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$. We denote by $N_{i,t}$ the number of times arm i is pulled by the end of round t, and let $\Delta_i = \mu^* - \mu_i$ be the gap between the expected reward of the optimal arm and the expected reward of arm i.

We now present the *Reward-Opportunity MAB* (R-O MAB) model. An instance of R-O MAB is represented by a tuple $\langle K, T, \boldsymbol{\mu}, f, \lambda \rangle$. The tuple $\langle K, T, \boldsymbol{\mu} \rangle$ is an instance of standard stochastic bandit as described above. The combination of f and λ creates what we call the "fairness policy".

The fairness requirements are expressed by a function f, $f : [0,1]^K \to [0,1]^K$. f receives as input a vector of expected rewards and outputs a vector of minimal fraction of times each arm has to be pulled in order not to be penalized. We let $f(\boldsymbol{\mu})_i$ denote the i'th entry of $f(\boldsymbol{\mu})$. We assume that $\sum_{i=1}^{K} f(\boldsymbol{\mu})_i \leq 1$ and that f is Lipschitz continuous with a Lipschitz constant L with respect to the l_1 norm. That is, for all $\boldsymbol{\mu}, \boldsymbol{\mu}' \in [0,1]^k$ it holds that $\|f(\boldsymbol{\mu}) - f(\boldsymbol{\mu}')\|_1 \leq L \|\boldsymbol{\mu} - \boldsymbol{\mu}'\|_1$. Satisfying the Lipschitz condition implies that two similar expected reward vectors get similar fairness requirements. From here on we call f the *fairness function*.

The difference between the fairness requirement and the number of times an arm was pulled, $Tf(\boldsymbol{\mu})_i - N_{i,T}$, represents the deviation from the fairness constraint. If the deviation is positive, it means the arm was not pulled enough times, i.e. the subpopulation did not receive enough opportunities according to the fairness function. In such a case, the decision-maker pays a cost. The paid cost for a single arm pull's deviation from the fairness requirement is given by λ_i, the *transfer cost* for arm i. If arm i was pulled less than

$Tf(\boldsymbol{\mu})_i$ times, the reward will be deducted by $\lambda_i(Tf(\boldsymbol{\mu})_i - N_{i,T})$. The transfer cost is known to the decision-maker in advance. To account for all cases, the possible cost which stems from the deviation is $\lambda_i \max\{Tf(\boldsymbol{\mu})_i - N_{i,T}, 0\}$. For simplicity, we use $\lambda_i = \lambda$ for all $i \in [K]$, but stress that our results also hold with minor modifications in the general case.

The utility of the decision-maker is denoted by $\mathcal{U}_{\lambda,f}$. It is an additive utility of the reward minus the total deviation from the fairness requirement. Notice that i_t and consequently $N_{i,T}$ depend on the algorithm playing the arms. Formally, given an algorithm ALG,

$$\mathcal{U}_{\lambda,f}(ALG;T) \stackrel{\text{def}}{=} \sum_{t=1}^{T} r_{i_t} - \lambda \sum_{i=1}^{k} \max\{Tf(\boldsymbol{\mu})_i - N_{i,T}, 0\}. \tag{1}$$

As is customary in the MAB literature, we focus on the *regret* of the decision-maker, which we denote $\mathcal{R}_{\lambda,f}(ALG;T)$. Let OPT be an algorithm maximizing the utility $\mathcal{U}_{\lambda,f}(OPT)$ (we discuss OPT in Subsection 2.1). The regret is the gap between the expected utility of OPT and ALG:

$$\mathcal{R}_{\lambda,f}(ALG;T) = \mathbb{E}(\mathcal{U}_{\lambda,f}(OPT;T)) - \mathbb{E}(\mathcal{U}_{\lambda,f}(ALG;T)). \tag{2}$$

When λ and f are arbitrary or clear from the context, we omit the subscript and simply denote $\mathcal{U}$ and $\mathcal{R}$. Full proofs appear in the appendix.

2.1 Optimal Algorithm

The structure of the optimal algorithm in classic MABs is straightforward: In every round, pick the arm with the highest expectation. However, in our case, the transfer cost makes the optimal algorithm a bit more complex, as we now elucidate. Let i denote an arbitrary index of a sub-optimal arm, i.e., an arm such that $\mu_i < \max_{i' \in [K]} \mu_{i'}$. The decision-maker has to decide whether to support the subpopulation associated with that arm explicitly (by pulling it $Tf(\boldsymbol{\mu})_i$ times) or implicitly (by paying $\lambda Tf(\boldsymbol{\mu})_i$). Note that $Tf(\boldsymbol{\mu})_i$ can be non-integer, in this case, we take the floor of $Tf(\boldsymbol{\mu})_i$. In each one of those $Tf(\boldsymbol{\mu})_i$ rounds, the decision-maker loses Δ_i if she pulls arm i (as she could pick the optimal arm) but saves λ (as she does not need the pay the transfer cost). Therefore, if the reward gap of arm i is greater than the transfer cost, $\Delta_i > \lambda$, the decision-maker does not pull arm i at all and pays the transfer cost. Otherwise, if $\Delta_i < \lambda$, the decision-maker would have greater utility by pulling arm i exactly $Tf(\boldsymbol{\mu})_i$ times and not incurring the transfer cost. If $\Delta_i = \lambda$, the decision-maker is indifferent between the two options. More formally,

Lemma 1. *Fix an arbitrary instance $\langle K, T, \boldsymbol{\mu}, f, \lambda \rangle$ and let OPT be an optimal algorithm for that instance. For every sub-optimal arm i, if $\Delta_i < \lambda$ then OPT pulls i exactly $Tf(\boldsymbol{\mu})_i$ times; if $\Delta_i > \lambda$, OPT does not pull i at all. If $\Delta_i = \lambda$, OPT pulls arm i between zero and $Tf(\boldsymbol{\mu})_i$ times.*

2.2 About the Fairness Policy

The fairness policy is comprised of the fairness function f and the transfer cost λ. f represents the decision-maker's view on how opportunities should be distributed. E.g., the zero function $f^0(\boldsymbol{\mu})_i \stackrel{\text{def}}{=} 0$ corresponds to standard Multi-Armed bandit problem without any constraints. Generalizing this case for any constant function, e.g., $f^{\text{uni}}(\boldsymbol{\mu})_i \stackrel{\text{def}}{=} \frac{1}{K}$, alludes that the decision-maker believes that all subpopulations are entitled to the same share of opportunities irrespective of their expected rewards. f can also grow linearly with each expected reward, for instance $f^{\text{lin}}(\boldsymbol{\mu})_i \stackrel{\text{def}}{=} \frac{\mu_i}{K}$. In the most general case, the number of required opportunities to a subpopulation can also depend on its expected reward relative to the expected rewards of other subpopulations, .e.g., $f^{\text{sft}}(\boldsymbol{\mu}; c)_i \stackrel{\text{def}}{=} \frac{\exp^{c\mu_i}}{\sum_{j=1}^{K} \exp^{c\mu_j}}$. Our modelling supports these special cases and many other natural candidates for the fairness function. Selecting λ complements the decision-maker's view on revenue and opportunities. As described in Section 2.1, if the transfer cost is high the decision-maker will tend to grant the opportunities explicitly, and would not grant opportunities explicitly only when the subpopulations' expected rewards have big differences and vice versa if the transfer cost is low. λ can vary between different subpopulations, for simplicity is assumed equal.

3 No-Regret Algorithms

In this section, we present our main algorithmic contribution. We devise *Self-regulated Utility Maximization*, which incurs a regret of $\tilde{O}(T^{2/3})$. Before we discuss it, we first demonstrate that classical MAB algorithms fail miserably on our setting. This is expected given that such algorithms were not devised for a setting like ours, but it will serve us later on. Classical MAB algorithms are tuned to pull sub-optimal arms as little as possible. As shown in Subsection 2.1, it is not always optimal for R-O MAB. If the cost of opportunity (Δ_i) is lower than the transfer cost (λ), the optimal algorithm pulls arm i according to the fairness function.

In R-O MAB, we face a unique challenge comparing to the classic MAB problem. Classical MAB algorithms are aimed at identifying the optimal arm but do not estimate the expected rewards $\boldsymbol{\mu}$. The optimal algorithm depends on the relation between the reward gaps and the transfer cost; hence, unlike classic MAB, accurate approximation of the reward gaps $(\Delta_i)_{i \in [K]}$ is crucial for our problem. Additionally, $f(\boldsymbol{\mu})$ should also be approximated correctly for arms i with $\Delta_i < \lambda$, to align with the optimal algorithm. These two challenges are singular to our settings and are reflected in the lower bound.

Algorithm 1, which we term Fairness-Aware-ETC, is a modified version of Explore-Then-Commit (ETC). ETC explores all arms for a predetermined number of rounds (N), and then follows the best preforming arm for the remaining rounds. Similarly, Fairness-Aware-ETC pulls each arm N times and then constructs estimates for $\boldsymbol{\mu}$ and $f(\boldsymbol{\mu})$, which we denote using the hat notation, i.e., $\hat{\boldsymbol{\mu}}$ and $f(\hat{\boldsymbol{\mu}})$. It then continues optimally with respect to these estimates (according to the optimal algorithm for the estimated quantities).

Theorem 1. *Fix any arbitrary instance of R-O MAB, and let $N = 8L^{2/3}T^{2/3}\log^{1/3} T$. Algorithm 1 has a regret of $O(KL^{2/3}T^{2/3}\log^{1/3} T)$.*

Algorithm 1 is almost data independent. The predefined exploration length N prevents the algorithm from stopping the exploration early. Early stopping is important after iden-

Algorithm 1 Fairness-Aware-ETC

Input: N - # exploration rounds

1: **for** $i = 1, \dots K$ **do**
2: pull arm i for N rounds
3: **end for**
4: **for** $i = 1, \dots K$ **do**
5: **if** $\hat{\Delta}_i < \lambda$ **then**
6: pull arm i for $\max\{Tf(\hat{\boldsymbol{\mu}})_i - N, 0\}$ rounds
7: **end if**
8: **end for**
9: pull an arbitrary arm from $\arg\max_{i\in[K]} \hat{\mu}_i$ until the execution ends

tifying arms with high opportunity cost or arms that already satisfy the fairness requirements.

3.1 Fairness Aware Black-Box algorithm

In this section, we present a data dependent algorithm addressing the problems of Algorithm 1 by incorporating early stopping. We now explain the course of Algorithm 2. Full version of the algorithm appears in the appendix. The algorithm takes α and β, which we describe shortly, and ALG, a black-box no-regret MAB algorithm as input, where ALG is no-regret with respect to the classical, rewards-only MAB objective (e.g. UCB1 [Slivkins, 2019]).

In Lines 1- 3 confidence bounds representing the probable estimates of the reward gaps, i.e. $LCB(\Delta_i)$ and $UCB(\Delta_i)$ and C_t which is the hyper-cube of probable estimates of $\boldsymbol{\mu}$ are initialized. Lines 5-13 consist of four different phases. In the first phase (Lines 4-5), the reward gaps are approximated. After this phase, the algorithm knows w.h.p. for each arm whether its reward gap is higher or lower than the transfer cost by more than β. The second phase (Lines 7-8), approximates f for arms with low opportunity cost up to a factor of α. If there is an arm with low opportunity cost for which the approximation of f is not accurate enough, all the arms are pulled. The term $\max_{\boldsymbol{\mu'}\in C_t} f(\boldsymbol{\mu'})_i - \min_{\boldsymbol{\mu'}\in C_t} f(\boldsymbol{\mu'})_i$ upper bounds estimation error of $f(\hat{\boldsymbol{\mu}})_i$ inside the hyper-cube C_t. Pulling all arms ensures that all the estimates improve for the subsequent round, namely, C_t shrinks in all of its dimensions. In the third phase (Lines 10-11), we ensure that we pull all arms with low opportunity cost according to the estimate of $f(\hat{\boldsymbol{\mu}})_i$. In the fourth step (Line 13), ALG is invoked until the end of the execution.

Next, we discuss the input hyper-parameters, α, β and ALG. α is the confidence interval hyper-parameter for the approximation of f. Setting α too small values implies that arms should be pulled many times and this can inflict a regret due to over pulling arms. The approximation error of f can be as big as $T\alpha$ is. The hyper-parameter β is the confidence interval for the approximation of the reward gaps. If the reward gap is not close to the transfer cost λ, it would be identified almost immediately. Otherwise, Algorithm 2 uses the black-box MAB algorithm ALG. In the fourth step, the decision-maker identifies the best arm and exploits its reward.

The only computationally non-trivial step in Algorithm 2 appears in Line 7: Computing $\max_{\boldsymbol{\mu'}\in C_t} f(\boldsymbol{\mu'})_i - \min_{\boldsymbol{\mu'}\in C_t} f(\boldsymbol{\mu'})_i$. Finding the global maximum of a Lipschitz function inside a hyper-cube is a computationally challenging task. However, due to role f plays in our setting, we argue that it should have a natural structure. Indeed, f quantifies a societal requirement and as such should be easy to grasp: Providing opportunities according to a cumbersome, hard-to-optimize and unexplainable criteria is likely to be unfair in and of itself. Consequentially, we shall assume that there is an oracle that computes the minimal and maximal values f at entry i can obtain in a given hyper-cube.

Algorithm 2 Self-regulated Utility Maximization

Input: Black-box bandit algorithm ALG, allowed approximation error parameters α and β

1: $N_i = 0, LCB(\Delta_i) = 0, UCB(\Delta_i) = 1$ for all $i \in [K]$
2: $t = 1$
3: $C_1 = [0, 1]^K$
4: **while** $\exists i \in [K]$ s.t $UCB(\Delta_i) > \lambda + \beta$ and $LCB(\Delta_i) < \lambda - \beta$ **do** // phase 1
5: Play all arms once, update t, counters and estimators
6: **end while**
7: **while** $\exists i \in [K]$ s.t. $\max_{\boldsymbol{\mu'}\in C_t} f(\boldsymbol{\mu'})_i - \min_{\boldsymbol{\mu'}\in C_t} f(\boldsymbol{\mu'})_i > \alpha$ and $LCB(\Delta_i) < \lambda$ **do** // phase 2
8: Play all arms once, update t, counters and estimators
9: **end while**
10: **while** $t < T$ and $\exists i \in [K]$ s.t $LCB(\Delta_i) < \lambda$ and $N_i < T \min_{\boldsymbol{\mu'}\in C_t} f(\boldsymbol{\mu'})_i$ **do** // phase 3
11: Play arm i the minimal number of times so $N_i \geq Tf(\hat{\boldsymbol{\mu}})_i$, update t
12: **end while**
13: Invoke ALG for the remaining rounds // phase 4

We are ready to state the guarantees of Algorithm 2.

Theorem 2. *Fix any arbitrary instance of R-O MAB, and let* $\alpha = K^{4/3}L^{2/3}T^{-1/3}\log^{1/3} T$, $\beta = T^{-1/3}\log^{1/3} T$. *Then, Algorithm 2 has a regret of* $O(K^{4/3}L^{2/3}T^{2/3}\log^{1/3} T)$.

4 Lower Bound

In the previous section, we presented Algorithm 2, which incurs a regret of $\tilde{O}\left(T^{2/3}\right)$ in the worst case. Here we show that this bound is asymptotically optimal by designing a family of R-O MAB instances that can mislead any algorithm.

Theorem 3. *Fix time horizon T, number of arms K, and Lipschitz constant L. For any algorithm, there exists a R-O MAB instance such that* $\mathcal{R}(T) \geq \Omega(T^{2/3})$.

5 Conclusion

We introduced a MAB problem that models decision making from the perspective of Corporate Social Responsibility and allocation of opportunities. Our modeling imitates many real-world scenarios where decision-makers are required to maximize their short-term utility while at the same time upholding fairness principles. With our framework, commercial companies can incorporate self-regulation in their algorithmic products, and provide opportunities as a form of social responsibility. We devised a no-regret algorithm and showed that its convergence rate is in fact optimal.

References

Amazon scraps secret AI recruiting tool that showed bias against women - reuters. URL https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

How to hire with algorithms. URL https://hbr.org/2016/10/how-to-hire-with-algorithms.

Using mobile to reach the Latin american unbanked — fico. URL https://www.fico.com/en/node/8140?file=7900.

F. Abel. We know where you should work next summer: Job recommendations. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 230–230, 2015.

A. Acharya and D. Sinha. Early prediction of students performance using machine learning techniques. *International Journal of Computer Applications*, 107(1), 2014.

S. Agrawal and N. Goyal. Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on learning theory*, pages 39–1, 2012.

I. Ajunwa and D. Greene. Platforms at work: Automated hiring platforms and other new intermediaries in the organization of work. *SP Vallas, and A. Kovalainen, Work and Labor in the Digital Age*, pages 61–91, 2019.

P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47 (2-3):235–256, 2002.

R. Berk. *Criminal justice forecasts of risk: A machine learning approach*. Springer Science & Business Media, 2012.

R. Berk, R. Berk, and Drougas. *Machine learning risk assessments in criminal justice settings*. Springer, 2019.

A. B. Carroll et al. The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders. *Business horizons*, 34(4):39–48, 1991.

S.-C. Chow and M. Chang. Adaptive design methods in clinical trials–a review. *Orphanet journal of rare diseases*, 3 (1):11, 2008.

A. B. Diekman, E. R. Brown, A. M. Johnston, and E. K. Clark. Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological science*, 21(8):1051–1057, 2010.

C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226, 2012.

M. C. Fu. Alphago and monte carlo tree search: the simulation optimization perspective. In *Proceedings of the 2016 Winter Simulation Conference*, pages 659–670. IEEE Press, 2016.

A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. Predictably unequal? the effects of machine learning on credit markets. *The Effects of Machine Learning on Credit Markets (November 6, 2018)*, 2018.

E. Garriga and D. Melé. Corporate social responsibility theories: Mapping the territory. *Journal of business ethics*, 53 (1-2):51–71, 2004.

M. Hardt, E. Price, N. Srebro, et al. Equality of opportunity in supervised learning. In *Advances in neural information processing systems*, pages 3315–3323, 2016.

M. Joseph, M. Kearns, J. Morgenstern, S. Neel, and A. Roth. Fair algorithms for infinite and contextual bandits. *arXiv preprint arXiv:1610.09559*, 2016a.

M. Joseph, M. Kearns, J. H. Morgenstern, and A. Roth. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*, pages 325–333, 2016b.

A. Lambrecht and C. E. Tucker. Algorithmic bias? an empirical study into apparent gender-based discrimination in the display of stem career ads. *An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads (March 9, 2018)*, 2018.

Y. Liu, G. Radanovic, C. Dimitrakakis, D. Mandal, and D. C. Parkes. Calibrated fairness in bandits. *arXiv preprint arXiv:1707.01875*, 2017.

T. Lux, R. Pittman, M. Shende, and A. Shende. Applications of supervised learning techniques on undergraduate admissions data. In *Proceedings of the ACM International Conference on Computing Frontiers*, pages 412–417, 2016.

A. Mas-Colell, M. D. Whinston, J. R. Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.

H. B. McMahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin, et al. Ad click prediction: a view from the trenches. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1222–1230, 2013.

Northpointe. Practitioner's guide to compas core, 2015. URL https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf.

R. Oentaryo, E.-P. Lim, M. Finegold, D. Lo, F. Zhu, C. Phua, E.-Y. Cheu, G.-E. Yap, K. Sim, M. N. Nguyen, et al. Detecting click fraud in online advertising: a data mining approach. *The Journal of Machine Learning Research*, 15(1): 99–140, 2014.

V. Patil, G. Ghalme, V. Nair, and Y. Narahari. Achieving fairness in the stochastic multi-armed bandit problem. *arXiv preprint arXiv:1907.10516*, 2019.

A. Pérez-Martín, A. Pérez-Torregrosa, and M. Vaca. Big data techniques to measure credit banking risk in home equity loans. *Journal of Business Research*, 89:448–454, 2018.

A. Slivkins. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019. ISSN 1935-8237. doi: 10.1561/2200000068. URL http://dx.doi.org/10.1561/2200000068.

H. R. Varian and H. R. Varian. *Microeconomic analysis*, volume 3. Norton New York, 1992.

A. Waters and R. Miikkulainen. Grade: Machine learning support for graduate admissions. *AI Magazine*, 35(1):64–64, 2014.

J. White. *Bandit algorithms for website optimization*. " O'Reilly Media, Inc.", 2012.

M. B. Zafar, I. Valera, M. G. Rodriguez, and K. P. Gummadi. Fairness constraints: Mechanisms for fair classification. *arXiv preprint arXiv:1507.05259*, 2015.

C. Zeng, Q. Wang, S. Mokhtari, and T. Li. Online context-aware recommendation with time varying multi-armed bandit. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 2025–2034, 2016.

S. Zhang, W. Xiong, W. Ni, and X. Li. Value of big data to finance: observations on an internet credit service company in china. *Financial Innovation*, 1(1):17, 2015.

J. Zhao, T. Wang, M. Yatskar, V. Ordonez, and K.-W. Chang. Men also like shopping: Reducing gender bias amplification using corpus-level constraints. *arXiv preprint arXiv:1707.09457*, 2017.