Characterizing Truthful Multi-Armed Bandit Mechanisms

[Extended Abstract][*]

Moshe Babaioff
Microsoft Research
Mountain View, CA 94043
moshe@microsoft.com

Yogeshwer Sharma[†]
Cornell University
Ithaca, NY 14853
yogi@cs.cornell.edu

Aleksandrs Slivkins
Microsoft Research
Mountain View, CA 94043
slivkins@microsoft.com

ABSTRACT

We consider a multi-round auction setting motivated by pay-per-click auctions for Internet advertising. In each round the auctioneer selects an advertiser and shows her ad, which is then either clicked or not. An advertiser derives value from clicks; the value of a click is her private information. Initially, neither the auctioneer nor the advertisers have any information about the likelihood of clicks on the advertisements. The auctioneer's goal is to design a (dominant strategies) truthful mechanism that (approximately) maximizes the social welfare.

If the advertisers bid their true private values, our problem is equivalent to the *multi-armed bandit problem*, and thus can be viewed as a strategic version of the latter. In particular, for both problems the quality of an algorithm can be characterized by *regret*, the difference in social welfare between the algorithm and the benchmark which always selects the same "best" advertisement. We investigate how the design of multi-armed bandit algorithms is affected by the restriction that the resulting mechanism must be truthful. We find that truthful mechanisms have certain strong structural properties – essentially, they must separate exploration from exploitation – *and* they incur much higher regret than the optimal multi-armed bandit algorithms. Moreover, we provide a truthful mechanism which (essentially) matches our lower bound on regret.

Categories and Subject Descriptors

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems; K.4.4 [**Computers and Society**]: Electronic Commerce; F.1.2 [**Computation by Abstract Devices**]: Modes of Computation—*Online computation*; J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

theory, algorithms, economics

Keywords

mechanism design, truthful mechanisms, single-parameter auctions, multi-armed bandits, online learning

1. INTRODUCTION

In recent years there has been much interest in understanding the implication of strategic behavior on the performance of algorithms whose input is distributed among selfish agents. This study was mainly motivated by the Internet, the main arena of large scale interaction of agents with conflicting goals. The field of Algorithmic Mechanism Design [32] studies the design of mechanisms in computational settings (for background see the recent book [33] and survey [35]).

Much attention has been drawn to the market for sponsored search (e.g. [25, 17, 36, 29, 3]), a billions dollar market with numerous auctions running every second. Research on sponsored search mostly focus on equilibria of the Generalized Second Price (GSP) auction [17, 36], the auction that is most commonly used in practice (e.g. by Google and Yahoo), or on the design of truthful auctions [2]. All these auctions rely on knowing the rates at which users click on the different advertisements (a.k.a. Click-Through-Rates, or CTRs), and do not consider the process in which these CTRs are learned or refined over time by observing users' behavior. We argue that strategic agents would take this process into account, as it influences their utility. Prior work [20] focused on the implication of click fraud on the methods used to learn CTRs. We on the other hand are interested in the implications of the *strategic bidding* by the agents. Thus, we consider the problem of designing truthful sponsored search auctions when the process of learning the CTRs is a part of the game.

We are mainly interested in the interplay between the online learning and the strategic aspects of the problem. To isolate this issue, we consider the following setting, which is a natural *strategic* version of the multi-armed bandit (MAB) problem. In this setting, there are k agents. Each agent i has a single advertisement, and a *private* value $v_i > 0$ for every click she gets. The mechanism is an online algorithm that first solicits bids from the agents, and then runs for T rounds. In each round the mechanism picks an agent (using the bids and the clicks observed in the past rounds), displays her advertisement, and receives a feedback – if there

[*] The full version of this paper [8] is available from `http://arxiv.org/abs/0812.2291` .

[†] This research was done while the author was an intern at Microsoft Research, Silicon Valley Center.

EC'09, July 6–10, 2009, Stanford, California, USA.

was a click or not. Payments are assigned after round T. Each agent tries to maximize her own utility: the difference between the value that she derives from clicks and the payment she pays. We assume that initially no information is known about the likelihood of each agent to be clicked, and in particular there are no Bayesian priors.

We are interested in designing mechanisms which are truthful (in dominant strategies): every agent maximizes her utility by bidding truthfully, for any bids of the others and *for any clicks* that would have been received. The goal is to maximize the social welfare.[1] Since the payments cancel out, this is equivalent to maximizing the total value derived from clicks, where an agent's contribution to that total is her private value times the number of clicks she receives. We call this setting the *MAB mechanism design problem.*

In the absence of strategic behavior this problem reduces to a standard MAB formulation in which an algorithm repeatedly chooses one of the k alternatives ("arms") and observes the associated payoff: the value-per-click of the corresponding ad if the ad is clicked, and 0 otherwise. The crucial aspect in MAB problems is the tradeoff between acquiring more information (*exploration*) and using the current information to choose a good agent (*exploitation*). MAB problems have been studied intensively for the past three decades (see [12, 13, 18]). In particular, the above formulation is well-understood [6, 7, 14] in terms of *regret* relative to the benchmark which always chooses the same "best" alternative. This notion of regret naturally extends to the strategic setting outlined above, the total payoff being exactly equal to the social welfare, and the regret being exactly the loss in social welfare. Thus one can directly compare MAB algorithms and MAB mechanisms in terms of welfare loss (regret).

Broadly, we ask how the design of MAB algorithms is affected by the restriction of truthfulness: what is the difference between the best *algorithms* and the best *truthful mechanisms*? We are interested both in terms of the structural properties and the gap in performance (in terms of regret). We are not aware of any prior work that characterizes truthful learning algorithms or proves negative results on their performance.

Our contributions. We present two main contributions. First, we present a characterization of (dominant-strategy) truthful mechanisms. Second, we present a lower bound on the regret that such mechanisms must suffer. This regret is significantly larger than the regret of the best MAB algorithms.

Formally, a mechanism for the MAB mechanism design problem is a pair $(\mathcal{A}, \mathcal{P})$, where $\mathcal{A}$ is the *allocation rule* (essentially, an MAB algorithm), and $\mathcal{P}$ is the *payment rule.* Note that regret is completely determined by the allocation rule. As is standard in the literature, we focus on mechanisms in which each agent's payment (averaged over clicks) is between 0 and her bid; such mechanisms are called *normalized*, and they satisfy voluntary participation.

The setting we study is a *single-parameter auction*, the most studied and well-understood type of auctions. For such settings truthful mechanisms are fully characterized [30, 4]: a mechanism is truthful if and only if the allocation rule is monotone (by increasing her bid an agent cannot cause a decrease in the number of clicks she gets), and the payment rule is defined in a specific and (essentially) unique way. Yet, this characterization is *not* the right characterization for the MAB setting! The main problem is that in our setting click information for any agent that is not chosen at a given round is not available to the mechanism, and thus cannot be used in the computation of payments. Thus, the payment cannot depend on any unobserved clicks. We show that this has severe implications on the structure of truthful mechanisms.

The first notable property of a truthful mechanism is a much stronger version of monotonicity:

Definition 1.1. *A* realization *consists of click information for all agents at all rounds (including unobserved ones). An allocation rule is* pointwise monotone *if for each realization, each bid profile and each round, if an agent is played at the round, then she is also played after increasing her bid (fixing everything else).*

Let us consider allocation rules that satisfy the following two natural conditions. First, an allocation rule is *scale-free* if it is invariant under multiplying all bids by the same positive number (essentially, changing the currency unit). Second, it is *Independent of Irrelevant Alternatives* (*IIA*, for short) if for any given realization, bid profile and round, a change of bid of agent i cannot transfer the allocation in this round from agent j to agent l, where these are three distinct agents. (Note that the second condition trivially holds if there are only two agents.)

We show that any truthful mechanism must have a strict separation between exploration and exploitation. A crucial feature of exploration is the ability to influence the allocation in forthcoming rounds. To make this point more concrete, we call a round *influential* for a given realization if for some bid profile changing the realization for this round can affect the allocation in some future round. We show that in any such round, the allocation can not depend on the bids. Thus, influential rounds are essentially useless for exploitation.

Definition 1.2. *An allocation rule $\mathcal{A}$ is called* exploration-separated *if for any given realization, the allocation in any influential round for that realization does not depend on the bids.*

We are now ready to present our main structural result, which is in fact a complete characterization.

Theorem 1.3. *Consider the MAB mechanism design problem. Let $\mathcal{A}$ be a non-degenerate[2] deterministic allocation rule which is scale-free and satisfies IIA. Then mechanism $(\mathcal{A}, \mathcal{P})$ is normalized and truthful for some payment rule*

[1] Social welfare includes both the actioneer's revenue and the agents' utility. Since in practice different sponsored search platforms compete against one another, taking into account the agents' utility increases the platform's attractiveness to the advertisers.

[2] Non-degeneracy is a mild technical assumption, formally defined in "preliminaries", which ensures that (essentially) if a given allocation happens for some bid profile (b_i, b_{-i}) then the same allocation happens for all bid profiles (x, b_{-i}), where x ranges over some non-degenerate interval. Without this assumption, all structural results hold (essentially) *almost surely* w.r.t the k-dimensional Lebesgue measure on the bid vectors. Exposition becomes significantly more cumbersome, yet leads to the same lower bounds on regret. For clarity, we assume non-degeneracy throughout this version of the paper.

$\mathcal{P}$ if and only if $\mathcal{A}$ is pointwise monotone and exploration-separated.

We also obtain a similar (but somewhat more complicated) characterization without assuming that allocations are scale-free and satisfy IIA (Theorem 3.8). We use it then to derive Theorem 1.3. We emphasize that our characterization results hold regardless of whether the auctioneer's goal is to maximize welfare or revenue or any other objective.

In view of Theorem 1.3, we present a lower bound on the performance of exploration-separated algorithms. We consider a setting, termed the *stochastic MAB mechanism design problem*, in which each click on a given advertisement is an independent random event which happens with a fixed probability, a.k.a. the CTR. The expected "payoff" from choosing a given agent is her private value times her CTR. For the ease of exposition, assume that the bids lie in the interval $[0, 1]$. Then the non-strategic version is the *stochastic MAB problem* in which the payoff from choosing a given arm i is an independent sample in $[0, 1]$ with a fixed mean μ_i. In both versions, *regret* is defined with respect to a hypothetical allocation rule (resp. algorithm) that always chooses an arm with the maximal expected payoff. Specifically, regret is the expected difference between the social welfare (resp. total payoff) of the benchmark and that of the allocation rule (resp. algorithm). The goal is to minimize $R(T)$, worst-case regret over all problem instances on T rounds.

We show that the worst-case regret of any exploration-separated mechanism is *larger* than that of the optimal MAB algorithm: $\Omega(T^{2/3})$ vs $O(\sqrt{T \log T})$ for a fixed number of agents. We obtain an even more pronounced difference if we restrict our attention to the *δ-gap* problem instances: instances for which the best agent is better than the second-best by a (comparatively large) amount δ, that is $\mu_1 v_1 - \mu_2 v_2 = \delta \cdot (\max_i v_i)$, where arms are arranged such that $\mu_1 v_1 \geq \mu_2 v_2 \geq \cdots \geq \mu_k v_k$. Such instances are known to be easy for the MAB algorithms. Namely, an algorithm can achieve the optimal worst-case regret $O(\sqrt{kT \log T})$ *and* regret $O(\frac{k}{\delta} \log T)$ on δ-gap instances [26, 6]. However, for exploration-separated mechanisms the worst-case regret $R_\delta(T)$ over the δ-gap instances is polynomial in T as long as worst-case regret is even remotely non-trivial (i.e., sublinear). Thus, for the δ-gap instances the gap between algorithms and truthful mechanisms in the worst-case regret is *exponential* in T.

Theorem 1.4. *Consider the stochastic MAB mechanism design problem with k agents. Let $\mathcal{A}$ be a deterministic allocation rule that is exploration-separated. Then $\mathcal{A}$ has worst-case regret $R(T) = \Omega(k^{1/3}\, T^{2/3})$. Moreover, if $R(T) = O(T^\gamma)$ for some $\gamma < 1$ then for every fixed $\delta \leq \frac{1}{4}$ and $\lambda < 2(1-\gamma)$ the worst-case regret over the δ-gap instances is $R_\delta(T) = \Omega(\delta\, T^\lambda)$.*

We note that our lower bounds holds for a more general setting in which the values-per-click can change over time, and the advertisers are allowed to change their bids at every time step.

To complete the picture, we present a very simple (deterministic) mechanism that is truthful and normalized, and matches the lower bound $R(T) = \Omega(k^{1/3}\, T^{2/3})$ up to logarithmic factors.

We also provide a number of extensions. First, we prove a similar (but slightly weaker) regret bound without the scale-free assumption. Second, we extend some of our results to randomized mechanisms; in this setting, (dominant-strategy) truthfulness means "truthfulness for each realization of the private randomness". Third, we consider a weaker notion of truthfulness for randomized mechanisms – for each realization of the clicks, but in expectation over the random seed, and use this notion to provide algorithmic results for the version of the MAB mechanism design problem in which clicks are chosen by an adversary. Fourth, we discuss an even more permissive notion of truthfulness – truthfulness in expectation over the clicks.

Other related work and discussion. The question of how the performance of a truthful mechanism compares to that of the optimal algorithm for the corresponding non-strategic problem has been considered in the literature in a number of other auction settings. Performance gaps have been shown for various scheduling problems [4, 32, 16] and for online auction for expiring goods [28]. Other papers presented approximation gaps due to *computational constraints*, e.g. for combinatorial auctions [27, 16] and combinatorial public projects [34], showing a gap via a structural result for truthful mechanisms.

The study of MAB mechanisms has been initiated by Gonen and Pavlov [19]. The authors present a MAB mechanism which is claimed to be truthful in a certain approximate sense. Unfortunately, this mechanism does not satisfy the claimed properties; this was also confirmed with the authors through personal communication (see also a similar note in [15]).

MAB algorithms were used in the design of Cost-Per-Action sponsored search auctions in Nazerzadeh et al. [31], where the authors construct a mechanism with approximate properties of truthfulness and individual rationality. Approximately truthful mechanisms are reasonable assuming the agents would not lie unless it leads to significant gains. However, this solution concept is weaker than the exact notion and it may still be rational for the agents to deviate (perhaps significantly) from being truthful. Moreover, as truthful bidding is not a Nash equilibrium, agents might have an increased incentive to deviate if they speculate that others are deviating. All of that may result in unpredictable, and possibly highly suboptimal outcomes. In this paper we focus on understanding what can be achieved with the *exact* truthfulness, mainly proving results of structural and lower-bounding nature. We note in passing that providing similar results for the approximately truthful setting such as the one in [31] is a worthy and challenging open question.

Independently and concurrently, Devanur and Kakade [15] have studied truthful MAB mechanisms with focus on maximizing the revenue. They present a lower bound of $\Omega(T^{2/3})$ on the loss in revenue with respect to the VCG (Vickrey-Clarke-Groves) payment, as well as a truthful mechanism that matches the lower bound. (This mechanism is almost identical to the one that we present in order to match the lower bound in Theorem 4.1.)

Our lower bounds use (a novel application of) the relative entropy technique from [26, 7], see [23] for an account. For other application of this technique, see e.g. [14, 21, 24, 10].

Our work focuses on regret in a prior-free setting in which the algorithm has no prior on CTRs. This is in contrast to the recent line of work on *dynamic auctions* [11, 5] which considers fully Bayesian settings in which there is a known prior on CTRs, and VCG-like social welfare-maximizing mechanisms are feasible. In our prior-free setting VCG-mechanisms cannot be applied as such mechanisms require the allocation to exactly maximize the expected social welfare, which is impossible (and not well-defined) without a prior.

We require the mechanisms to satisfy a strong notion of truthfulness: bidding truthful is optimal for *every* possible realization. This is desirable as it does not require the agents to be risk neutral. Moreover, such notion does not require agents to consider the process that generates the clicks. In particular, even in the presence of click spamming by others an agent's best strategy is still to bid truthfully. Finally, an agent never regrets in retrospect that she has been truthful.

Map of the paper. Section 2 is preliminaries. Truthfulness characterization is developed and proved in Section 3. The lower bounds on regret and the simple mechanism that matches them are in Section 4. Extensions and open questions are in Section 5. Due to the page limit, some of the proofs are deferred to the full version [8]

2. DEFINITIONS AND PRELIMINARIES

In the MAB mechanism design problem, there is a set K of k agents numbered from 1 to k. Each agent i has a *value* $v_i > 0$ for every click she gets; this value is known only to agent i. Initially, each agent i submits a *bid* $b_i > 0$, possibly different from v_i. [3] The "game" lasts for T rounds, where T is the given *time horizon*. A *realization* represents the click information for all agents and all rounds. Formally, it is a tuple $\rho = (\rho_1\,,\ \ldots\ ,\rho_k)$ such that for every agent i and round t, the bit $\rho_i(t) \in \{0, 1\}$ indicates whether i gets a click if played at round t. An *instance* of the MAB mechanism design problem consists of the number of agents k, time horizon T, a vector of private values $v = (v_1, \ldots, v_k)$, a vector of bids (*bid profile*) $b = (b_1, \ldots, b_k)$, and realization ρ.

A *mechanism* is a pair $(\mathcal{A}, \mathcal{P})$, where $\mathcal{A}$ is allocation rule and $\mathcal{P}$ is the payment rule. An *allocation rule* is represented by a function $\mathcal{A}$ that maps bid profile b, realization ρ and a round t to the agent i that is chosen (receives an *impression*) in this round: $\mathcal{A}(b; \rho; t) = i$. We also denote $\mathcal{A}_i(b; \rho; t) = \mathbf{1}_{\{\mathcal{A}(b;\rho;t)=i\}}$. The allocation is *online* in the sense that at each round it can only depend on clicks observed prior to that round. Moreover, it does not know the realization in advance; in every round it only observes the realization for the agent that is shown in that round. A *payment rule* is a tuple $\mathcal{P} = (\mathcal{P}_1\,,\ \ldots\ ,\mathcal{P}_k)$, where $\mathcal{P}_i(b; \rho) \in \mathbb{R}$ denotes the payment charged to agent i when the bids are b and the realization is ρ. [4] The payment can only depend on observed clicks. A mechanism is called *normalized* if for any agent i, bids b and realization ρ it holds that $\mathcal{P}_i(b; \rho)$ is non-negative and at most b_i times the number of clicks agent i got.

For given realization ρ and bid profile b, the number of clicks received by agent i is denoted $\mathcal{C}_i(b; \rho)$. Call $\mathcal{C} = (\mathcal{C}_1\,,\ \ldots\ ,\mathcal{C}_k)$ the *click-allocation* for $\mathcal{A}$. The *utility* that agent i with value v_i gets from the mechanism $(\mathcal{A}, \mathcal{P})$ when the bids are b and the realization is ρ is $\mathcal{U}_i(v_i; b; \rho) = v_i \cdot \mathcal{C}_i(b; \rho) - \mathcal{P}_i(b; \rho)$ (quasi-linear utility). The mechanism is *truthful* if for any agent i, value v_i, bid profile b and realization ρ it is the case that $\mathcal{U}_i(v_i; v_i, b_{-i}; \rho) \geq \mathcal{U}_i(v_i; b_i, b_{-i}; \rho)$.

In the *stochastic* MAB mechanism design problem, an adversary specifies a vector $\mu = (\mu_1\,,\ \ldots\ ,\mu_k)$ of CTRs (concealed from $\mathcal{A}$), then for each agent i and round t, realization $\rho_i(t)$ is chosen independently with mean μ_i. Thus, an instance of the problem includes μ rather than a fixed realization. For a given problem instance $\mathcal{I}$, let $i^* \in \text{argmax}_i\, \mu_i\, v_i$, then *regret* on this instance is defined as

$$R^{\mathcal{I}}(T) = T\, v_{i^*}\, \mu_{i^*} - \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k} \mu_i\, v_i\, \mathcal{A}_i(b;\, \rho;\, t)\right]. \quad (2.1)$$

For a given parameter $v_{\max}$, the *worst-case regret*[5] $R(T; v_{\max})$ denotes the supremum of $R^{\mathcal{I}}(T)$ over all problem instances $\mathcal{I}$ in which all private values are at most $v_{\max}$. Similarly, we define $R_\delta(T; v_{\max})$, the *worst-case δ-regret*, by taking the supremum only on instances with δ-gap.

Most of our results are stated for *non-degenerate* allocation rules, defined as follows. An interval is called *non-degenerate* if it has positive length. Fix bid profile b, realization ρ, and rounds t and t' with $t \leq t'$. Let $i = \mathcal{A}(b; \rho; t)$ and ρ' be the allocation obtained from ρ by flipping the bit $\rho_i(t)$. An allocation rule $\mathcal{A}$ is *non-degenerate* w.r.t. (b, ρ, t, t') if there exists a non-degenerate interval $I \ni b_i$ such that

- $\mathcal{A}_i(x, b_{-i}; \varphi; s) = \mathcal{A}_i(b; \varphi; s)$
- for each $\varphi \in \{\rho, \rho'\}$, each $s \in \{t, t'\}$, and all $x \in I$.

An allocation rule is *non-degenerate* if it is non-degenerate w.r.t. each tuple (b, ρ, t, t').

3. TRUTHFULNESS CHARACTERIZATION

Before presenting our characterization we begin by describing some related background. The click allocation $\mathcal{C}$ is *non-decreasing* if for each agent i, increasing her bid (and keeping everything else fixed) does not decrease $\mathcal{C}_i$. Prior work has established a characterization of truthful mechanisms for single-parameter domains (domains in which the private information of each agent is one-dimensional), relating click allocation monotonicity and truthfulness (see below). For our problem, this result is a characterization of MAB algorithms that are truthful for a given realization ρ, assuming that the *entire* realization ρ can be used to compute payments (when computing payments one can use click information for every round and every agent, even if the agent was not shown at that round.) One of our main contributions is a characterization of MAB allocation rules that can be truthfully implemented when payment computation is restricted to only use clicks information of the actual impressions assigned by the allocation rule.

An MAB allocation rule $\mathcal{A}$ is *truthful with unrestricted payment computation* if it is truthful with a payment rule

[3] One can also consider a more realistic and general model in which the value-per-click of an agent changes over time and the agents are allowed to change their bid at every round. The case that the value-per-click of each agent does not change over time is a special case. In that case truthfulness implies that each agent basically submits one bid as in our model (the same bid at every round), thus our main results (necessary conditions for truthfulness and regret lower bounds) also hold for the more general model.

[4] We allow the mechanism to determine the payments at the end of the T rounds, and not after every round. This makes that task of designing a truthful mechanism *easier* and thus strengthen our necessary condition for truthfulness (the condition used to derive the lower bounds on regret.)

[5] By abuse of notation, when clear from the context, the "worst-case regret" is sometimes simply called "regret".

that can use the *entire* realization ρ in it computation. We next present the prior result characterizing truthful mechanisms with unrestricted payment computation.

THEOREM 3.1 (MYERSON [30], ARCHER AND TARDOS [4]). *Let $(\mathcal{A}, \mathcal{P})$ be a normalized mechanism for the MAB mechanism design problem. It is truthful with unrestricted payment computation if and only if for any given realization ρ the corresponding click-allocation $\mathcal{C}$ is non-decreasing and the payment rule is given by*[6]

$$\mathcal{P}_i(b_i, b_{-i}; \rho) = b_i \cdot \mathcal{C}_i(b_i, b_{-i}; \rho) - \int_0^{b_i} \mathcal{C}_i(x, b_{-i}; \rho)\, dx. \quad (3.1)$$

We can now move to characterize truthful MAB mechanisms when the payment computation is restricted. The following notation will be useful: for a given realization ρ, let $\rho \oplus \mathbf{1}(i,t)$, be the realization that coincides with ρ everywhere, except that the bit $\rho_i(t)$ is flipped.

The first notable property of truthful mechanisms is a stronger version of monotonicity. Recall (see Definition 1.1) that an allocation rule $\mathcal{A}$ is *pointwise monotone* if for each realization ρ, bid profile b, round t and agent i, if $\mathcal{A}_i(b_i, b_{-i}; \rho; t) = 1$ then $\mathcal{A}_i(b_i^+, b_{-i}; \rho; t) = 1$ for any $b_i^+ > b_i$. In words, increasing a bid cannot cause a loss of an impression.

LEMMA 3.2. *Consider the MAB mechanism design problem. Let $(\mathcal{A}, \mathcal{P})$ be a normalized truthful mechanism such that $\mathcal{A}$ is a non-degenerate deterministic allocation rule. Then $\mathcal{A}$ is pointwise-monotone.*

PROOF. For a contradiction, assume not. Then there is a realization ρ, a bid profile b, a round t and agent i such that agent i loses an impression in round t by increasing her bid from b_i to some larger value b_i^+. In other words, we have $\mathcal{A}_i(b_i^+, b_{-i}; \rho; t) < \mathcal{A}_i(b_i, b_{-i}; \rho; t)$. Without loss of generality, let us assume that there are no clicks after round t, that is $\rho_j(t') = 0$ for any agent j and any round $t' > t$ (since changes in ρ after round t does not affect anything before round t).

Let $\rho' = \rho \oplus \mathbf{1}(i,t)$. The allocation in round t cannot depend on this bit, so it must be the same for both realizations. Now, for each realization $\varphi \in \{\rho, \rho'\}$ the mechanism must be able to compute the price for agent i when bids are (b_i^+, b_{-i}). That involves computing the integral $I_i(\varphi) = \int_{x \le b_i^+} \mathcal{C}_i(x, b_{-i}; \varphi)\, dx$ from (3.1). We claim that $I_i(\rho) \neq I_i(\rho')$. However, the mechanism cannot distinguish between ρ and ρ' since they only differ in bit (i,t) and agent i does not get an impression in round t. This is a contradiction.

It remains to prove the claim. Without loss of generality, assume that $\rho_i(t) = 0$ (otherwise interchange the role of ρ and ρ'). We first note that $\mathcal{C}_i(x, b_{-i}; \rho) \le \mathcal{C}_i(x, b_{-i}; \rho')$ for every x. This is because everything is same in ρ and ρ' until round t (so the impressions are same too), there are no clicks after round t, and in round t the behavior of $\mathcal{A}$ on the two realizations can be different only if that agent i gets an impression, in which case she is clicked under ρ' and not clicked under ρ.

Since $\mathcal{A}$ is non-degenerate, there exists a non-degenerate interval I containing b_i such that changing bid of agent i to any value in this interval does not change the allocation at round t (both for ρ and for ρ'). For any $x \in I$ we have $\mathcal{C}_i(x, b_{-i}; \rho) < \mathcal{C}_i(x, b_{-i}; \rho')$, where the difference is due to the click in round t. It follows that $I_i(\rho) < I_i(\rho')$. Claim proved. Hence, the mechanism cannot be implemented truthfully. □

Recall (see Definition 1.2) that round t is *influential* for a given realization ρ if for some bid profile b there exists a round $t' > t$ such that $\mathcal{A}(b; \rho; t') \neq \mathcal{A}(b; \rho \oplus \mathbf{1}(j,t); t')$ for $j = \mathcal{A}(b; \rho; t)$. In words: changing the relevant part of the realization at round t affects the allocation in some future round t'. An allocation rule $\mathcal{A}$ is called *exploration-separated* if for any given realization ρ and round t that is influential for ρ, it holds that $\mathcal{A}(b; \rho; t) = \mathcal{A}(b'; \rho; t)$ for any two bid vectors b, b' (allocation at t does not depend on the bids).

The main structural implication is "truthful implies exploration-separated". To illustrate the ideas behind this implication, we first state and prove it for two agents.

PROPOSITION 3.3. *Consider the MAB mechanism design problem with two agents. Let $\mathcal{A}$ be a non-degenerate scale-free deterministic allocation rule. If $(\mathcal{A}, \mathcal{P})$ is a normalized truthful mechanism for some $\mathcal{P}$, then it is exploration-separated.*

PROOF. Assume $\mathcal{A}$ is not exploration-separated. Then there is a *counterexample* (ρ, t): a realization ρ and a round t such that round t is influential and allocation in round t depends on bids. We want to prove that this leads to a contradiction.

Let us pick a counterexample (ρ, t) with some useful properties. Since round t is influential, there exists a realization ρ and bid profile b such that the allocation at some round $t' > t$ (the *influenced* round) is different under realization ρ and another realization $\rho' = \rho \oplus \mathbf{1}(j,t)$, where $j = \mathcal{A}(b; \rho; t)$ is the agent chosen at round t under ρ. Without loss of generality, let us pick a counterexample with minimum value of t' over all choices of (b, ρ, t). For ease of exposition, from this point on let us assume that $j = 2$. For the counterexample we can also assume that $\rho_1(t') = 1$, and that there are no clicks after round t', that is $\rho_l(t'') = \rho'_l(t'') = 0$ for all $t'' > t'$ and for all $l \in \{1, 2\}$.

We know that the allocation in round t depends on bids. This means that agent 1 gets an impression in round t for some bid profile $\hat{b} = (\hat{b}_1, \hat{b}_2)$ under realization ρ, that is $\mathcal{A}(\hat{b}; \rho; t) = 1$. As the mechanism is scale-free this means that, denoting $b_1^+ = \hat{b}_1\, b_2 / \hat{b}_2$ we have $\mathcal{A}(b_1^+, b_2; \rho; t) = 1$. Since $\mathcal{A}(b_1, b_2; \rho; t) = 2$ and $\mathcal{A}(b_1^+, b_2; \rho; t) = 1$, pointwise monotonicity (Lemma 3.2) implies that $b_1^+ > b_1$. We conclude that there exists a bid $b_1^+ > b_1$ for agent 1 such that $\mathcal{A}(b_1^+, b_2; \rho; t) = 1$.

Now, the mechanism needs to compute prices for agent 1 for bids $(b_1^+,\, b_2)$ under realizations ρ and ρ', that is $\mathcal{P}_1(b_1^+, b_2; \rho)$ and $\mathcal{P}_1(b_i^+, b_2; \rho')$. Therefore, the mechanism needs to compute the integral $I_1(\varphi) = \int_{x \le b_1^+} \mathcal{C}_1(x, b_2; \varphi)\, dx$ for both realizations $\varphi \in \{\rho, \rho'\}$.

First of all, for all $x \le b_1^+$ and for all $t'' < t'$, $\mathcal{A}(x, b_2; \rho; t'') = \mathcal{A}(x, b_2; \rho'; t'')$, since otherwise the minimality of t' will be violated. The only difference in the allocation can occur in round t'.

Let us assume $\mathcal{A}_1(b_1, b_2; \rho; t') < \mathcal{A}_1(b_1, b_2; \rho', t')$ (otherwise, we can swap ρ and ρ'). We make the claim that for all bids $x \le b_1^+$ of agent 1, the influence of round t on round t'

[6] Archer and Tardos [4] was the first paper in the Theoretical Computer Science literature that presented a characterization of truthful mechanisms for single-parameter domains, in the context of machine scheduling.

is in the same "direction":

$$\mathcal{A}_1(x, b_2; \rho; t') \leq \mathcal{A}_1(x, b_2; \rho'; t') \;\text{ for all }\; x \leq b_1^+. \qquad (3.2)$$

Suppose (3.2) does not hold. Then there is an $x < b_1^+$ such that $1 = \mathcal{A}_1(x, b_2; \rho; t') > \mathcal{A}_1(x, b_2; \rho'; t') = 0$. (Note that we have used the fact that the mechanism is deterministic.) If $x < b_1$ then pointwise monotonicity is violated under realization ρ, since $\mathcal{A}_1(x, b_2; \rho; t') > A_1(b_1, b_2; \rho; t')$; otherwise it is violated under realization ρ', giving a contradiction in both cases. The claim (3.2) follows.

Since $\mathcal{A}$ is non-degenerate, there exists a non-degenerate interval I containing b_i such that if agent 1 bids any value $x \in I$ then $\mathcal{A}_1(x, b_2; \rho; t') < \mathcal{A}_1(x, b_2; \rho'; t')$. Now by (3.2) it follows that $I_1(\rho) < I_2(\rho')$. However, the mechanism cannot distinguish between ρ and ρ' when the bid of agent 1 is b_1^+, since the differing bit $\rho_1(t)$ is not observed. Therefore the mechanism cannot compute prices, contradiction. □

3.1 General Truthfulness Characterization

Let us develop the general truthfulness characterization that does not assume that an allocation is scale-free and IIA. We will later use it to derive Theorem 1.3.

Definition 3.4. *Fix realization ρ and bid vector b. A round t is called* $(b;\rho)$-secured *from agent i if $\mathcal{A}(b_i^+, b_{-i}; \rho; t) = \mathcal{A}(b_i, b_{-i}; \rho; t)$ for any $b_i^+ > b_i$. A round t is called* bid-independent *w.r.t. ρ if the allocation $\mathcal{A}(b; \rho; t)$ is a constant function of b.*

The following definitions elaborate on the notion of an *influential round.*

Definition 3.5. *A round t is called* $(b;\rho)$-influential*, for bid profile b and realization ρ, if for some round $t' > t$ it holds that $\mathcal{A}(b; \rho; t') \neq \mathcal{A}(b; \rho'; t')$ for realization $\rho' = \rho \oplus \mathbf{1}(j, t)$ such that $j = \mathcal{A}(b; \rho; t)$.* [7] *In this case, t' is called the* influenced round *and j is called the* influencing agent *of round t. The agent i is called an* influenced agent *of round t if $i \in \{\mathcal{A}(b; \rho; t'),\, \mathcal{A}(b; \rho'; t')\}$.*

Note that a round is influential w.r.t. realization ρ if and only if it is (b, ρ)-influential for some b. The central property in our characterization is that each (b, ρ)-influential round is (b, ρ)-secured.

Definition 3.6. *A deterministic allocation is called* weakly separated *if for every realization ρ and each bid vector b, it holds that if round t is $(b;\rho)$-influential with influenced agent i then it is $(b;\rho)$-secured from i.*

We notice that exploration-separated is a stronger notion.

Observation 3.7. *For a deterministic allocation, exploration-separated implies weakly separated.*[8]

We are now ready to state our general characterization.

Theorem 3.8. *Consider the MAB mechanism design problem. Let $\mathcal{A}$ be a non-degenerate deterministic allocation rule. Then mechanism $(\mathcal{A}, \mathcal{P})$ is normalized and truthful for some payment rule $\mathcal{P}$ if and only if $\mathcal{A}$ is pointwise monotone and weakly separated.*

Proof. For the "only if" direction, $\mathcal{A}$ is pointwise-monotone by Lemma 3.2, and the fact that $\mathcal{A}$ is weakly separated is proved similarly to Proposition 3.3 (albeit with a few extra details). We defer it to the full version [8].

We focus on the "if" direction. Let $\mathcal{A}$ be a deterministic allocation rule which is pointwise monotone and weakly separated. We need to provide a payment rule $\mathcal{P}$ such that the resulting mechanism $(\mathcal{A}, \mathcal{P})$ is truthful and normalized. Since $\mathcal{A}$ is pointwise monotone, it immediately follows that it is monotone (i.e., as an agent increases her bid, the number of clicks that she gets cannot decrease). Therefore it follows from Theorem 3.1 that mechanism $(\mathcal{A}, \mathcal{P})$ is truthful and normalized if and only if $\mathcal{P}$ is given by (3.1). We need to show that $\mathcal{P}$ can be computed using only the knowledge of the clicks (bits from the realization) that were revealed during the execution of $\mathcal{A}$.

Assume we want to compute the payment for agent i in bid profile (b_i, b_{-i}) and realization ρ. We will prove that we can compute $\mathcal{C}_i(x) := \mathcal{C}_i(x, b_{-i}; \rho)$ for all $x \leq b_i$. To compute $\mathcal{C}_i(x)$, we show that it is possible to simulate the execution of the mechanism with $\texttt{bid}_i = x$. In some rounds, the agent i loses an impression, and in others it retains the impression (pointwise monotonicity ensures that agent i cannot gain an impression when decreasing her bid). In rounds that it loses an impression, the mechanism does not observe the bits of ρ in those rounds, so we prove that those bits are *irrelevant* while computing $\mathcal{C}_i(x)$. In other words, while running with $\texttt{bid}_i = x$, if mechanism needs to observe the bit that was not revealed when running with $\texttt{bid}_i = b_i$, we arbitrarily put that bit equal to 1 and simulate the execution of $\mathcal{A}$. We want to prove that this computes $\mathcal{C}_i(x)$ correctly.

Let $t_1 < t_2 < \cdots < t_n$ be the rounds in which agent i did not get an impression while bidding x, but did get an impression while bidding b_i. Let $\rho^0 := \rho$, and let us define realization ρ^l inductively for every $l \in [n]$ by setting $\rho^l := \rho^{l-1} \oplus \mathbf{1}(j_l, t_l)$, where $j_l = \mathcal{A}(x, b_{-i}; \rho^{l-1}; t_l)$ is the agent that got the impression at round t_l with realization ρ^{l-1} and bids (x, b_{-i}).

First, we claim that $j_l \neq i$ for any l. Indeed, suppose not, and pick the smallest l such that $j_{l+1} = i$. Then t_l is a $(x, b_{-i};\, \rho^l)$-influential round, with influenced agent $j_{l+1} = i$. Thus t_l is $(x, b_{-i};\, \rho^l)$-secured from i. Since $\mathcal{A}(x, b_{-i};\, \rho^l; t_l) = \mathcal{A}(x, b_{-i};\, \rho^{l-1}; t_l) = j_l \neq i$ by minimality of l, agent i does not get an impression in round t_l if she raises her bid to b_i. That is, $\mathcal{A}(b; \rho^l; t_l) \neq i$. However, the changes in realizations ρ^0 , ... , ρ^{l-1} only concern the rounds in which agent i is chosen, so they are not seen by the allocation if the bid profile is b (to prove this formally, use induction). Thus, $\mathcal{A}(b; \rho^l; t_l) = \mathcal{A}(b; \rho; t_l) = i$, contradiction. Claim proved. It follows that $\mathcal{A}(b; \rho; t_l) = i$ for each l. (This is because by induction, the change from ρ^{l-1} to ρ^l is not seen by the allocation if the bid profile is b.)

We claim that $\mathcal{A}_i(x, b_{-i}; \rho; t') = \mathcal{A}_i(x, b_{-i}; \rho^n; t')$ for every round t', which will prove the theorem. If not, then there exists l such that $\mathcal{A}_i(x, b_{-i}; \rho^l; t') \neq \mathcal{A}_i(x, b_{-i}; \rho^{l-1}; t')$ for some t' (and of course $t' > t_l$). Round t_l is thus $(x, b_{-i}; \rho^l)$-influential with influenced round t' and influenced agent i. Moreover, the influencing agent of that round is j_l, and we

[7] Note that realizations ρ and ρ' are interchangeable.

[8] To see this, simply use the definitions. Fix realization ρ and bid vector b, let t be a $(b;\rho)$-influential round with influenced agent i. We need to show that t is $(b;\rho)$-secured from i. Round t is $(b;\rho)$-influential, thus influential w.r.t. ρ, thus (since the allocation is exploration-separated) it is bid-independent w.r.t. ρ, thus agent i cannot change allocation in round t by increasing her bid.

already proved that $j_l \neq i$. Since round t_l is $(x, b_{-i}; \rho^l)$-secured from agent i due to the "weakly separated" condition, it follows that agent i does not get an impression in round t_l if she raises her bid to b_i. That is, $\mathcal{A}(b; \rho^l; t_l) \neq i$, contradiction. □

Note that we have proven the main characterization result (Theorem 1.3) for the case of two agents, because for two agents, it is not hard to see that a scale-free allocation is exploration-separated if and only if it is weakly separated, and also IIA trivially holds for two agents.

Let us argue that the non-degeneracy assumption in Theorem 3.8 is indeed necessary. To this end, let us present a simple deterministic mechanism $(\mathcal{A}, \mathcal{P})$ for two agents that is truthful and normalized, such that the allocation rule $\mathcal{A}$ is pointwise monotone, scale-free and yet *not* weakly separated. (The catch is, of course, that it is degenerate.) There are only two rounds. Agent 1 allocated at round 1 if and only if $b_1 \geq b_2$. Agent 1 allocated at round 2 if $b_1 > b_2$ or if $b_1 = b_2$ and $\rho_1(1) = 1$; otherwise agent 2 is shown. This completes the description of the allocation rule. To obtain a payment rule $\mathcal{P}$ which makes the mechanism normalized and truthful, consider an alternate allocation rule $\mathcal{A}'$ which in each round selects agent 1 if and only if $b_1 \geq b_2$. (Note that $\mathcal{A}' = \mathcal{A}$ except when $b_1 = b_2$.) Use Theorem 3.8 for $\mathcal{A}'$ to obtain a normalized truthful mechanism $(\mathcal{A}', \mathcal{P}')$, and set $\mathcal{P} = \mathcal{P}'$. The payment rule $\mathcal{P}$ is well-defined since the observed clicks for $\mathcal{P}$ and $\mathcal{P}'$ coincide unless $b_1 = b_2$, in which case both payment rules charge 0 to both players. The resulting mechanism $(\mathcal{A}, \mathcal{P})$ is normalized and truthful because the integral in (3.1) remains the same even if we change the value at a single point. It is easy to see that the allocation rule $\mathcal{A}$ has all the claimed properties; it fails to be non-degenerate because round t is influential only when $b_1 = b_2$.

3.2 Scalefree and IIA allocation rules

To complete the proof of Theorem 1.3, we show that under the right assumptions, an allocation is exploration-separated if and only if it is weakly separated. The full proof of this result is in the full version [8].

LEMMA 3.9. *Consider the MAB mechanism design problem. Let $\mathcal{A}$ be a non-degenerate deterministic allocation rule which is scalefree, pointwise monotone, and satisfies IIA. Then it is exploration-separated if and only if it is weakly separated.*

PROOF SKETCH. We sketch the proof of Lemma 3.9 at a *very* high level. The "only if" direction was observed in Observation 3.7. For the "if" direction, Let $\mathcal{A}$ be a weakly-separated mechanism. We prove by a contradiction that it is exploration-separated. If not, then there is a realization ρ and a round t such that t is influencial w.r.t. ρ as well as not bid-dependent w.r.t. ρ. Let round t be influencial with bid vector b, influencing agent l, and influenced agents j and $j' \neq j$ in influenced round t' (see $\boxed{1}$ in Figure 1; all boxed numbers in this sketch will refer to this figure).

From the assumption, t is not bid-dependent w.r.t. ρ, which means that there exists a bid profile b' such that $i' \neq l$ is played in round t with bids b'. Using scalefreeness, IIA, and pointwise-monotonicity, we can prove that there exists a sufficiently large bid $b^+_{i'}$ of agent i' such that she gets an impression in round t with bids $(b^+_{i'}, b_{-i'})$ (see $\boxed{2}$). Using the properties of the mechanism, it can further be proved that there is an agent i such that she gets the impression in round t when either i increases her bid, *or* l decreases her bid (see $\boxed{3}$). When i increases her bid to b^+_i, she also gets an impression in round t', since impressions cannot differ in round t' in the case when l is not played in round t and they must get transferred from j and j' to *somebody* in round t', and IIA implies that this *somebody* should be i.

Recall that two different players j and j' get the impression in round t' under ρ and ρ' respectively (see $\boxed{4}$). We prove that either agent j' or agent j must be equal to l (this is done by looking at how the allocation in round t' changes when l decreases her bid). Let us break the symmetry and assume $j' = l$ (see box $\boxed{5}$). It is also easy to see that when i increases her bid, impression in round t' get transferred to her in ρ (at some minimum value $b^{+\rho}_i$, see $\boxed{6}$), and impression in round t' gets transferred to her also in ρ' (as some possibly different minimum value $b^{+\rho'}_i$, see $\boxed{7}$). Using the assumptions of weakly-separatedness, we prove that $b^{+\rho}_i = b^{+\rho'}_i$ (see $\boxed{8}$). This can be proved by observing that $b^+_i \geq \max\{b^{+\rho}_i, b^{+\rho'}_i\}$, and then using weakly-separatedness of $\mathcal{A}$. Since these two bids were at a "threshold value" (these were the minimum values of bids to have transferred the impression in ρ and ρ' from j and l respectively), we are able to prove that the ratio of b_j/b_l must be some fixed number dependent on ρ, ρ', and t'. In particular, it follows that b_l belongs to a finite set $S(b_{-l})$ which depends only on b_{-l}. However, by non-degeneracy of $\mathcal{A}$ there must be infinitely many such b_l's, which leads to a contradiction. □

4. LOWER BOUNDS ON REGRET

In this section we use structural results from the previous section to derive lower bounds on regret.

THEOREM 4.1. *Consider the stochastic MAB mechanism design problem with k agents. Let $\mathcal{A}$ be an exploration-separated deterministic allocation rule. Then its regret is $R(T;\, v_{\max}) = \Omega(v_{\max}\, k^{1/3}\, T^{2/3})$.*

Let $\vec{\mu}_0 = (\frac{1}{2}, \ldots, \frac{1}{2}) \in [0,1]^k$ be the vector of CTRs in which for each agent the CTR is $\frac{1}{2}$. For each agent i, let $\vec{\mu}_i = (\mu_{i1}, \ldots, \mu_{ik}) \in [0,1]^k$ be the vector of CTRs in which agent i has CTR $\mu_{ii} = \frac{1}{2} + \epsilon$, $\epsilon = k^{1/3}\, T^{-1/3}$, and every other agent $j \neq i$ has CTR $\mu_{ij} = \frac{1}{2}$. As a notational convention, denote by $\mathbb{P}_i[\cdot]$ and $\mathbb{E}_i[\cdot]$ respectively the probability and expectation induced by the algorithm when clicks are given by $\vec{\mu}_i$. Let $\mathcal{I}_i$ be the problem instance in which CTRs are given by $\vec{\mu}_i$ and all bids are $v_{\max}$. For each agent i, let $\mathcal{J}_i$ be the problem instance in which CTRs are given by $\vec{\mu}_0$, the bid of agent i is $v_{\max}$, and the bids of all other agents are $v_{\max}/2$. We will show that for any exploration-separated deterministic allocation rule $\mathcal{A}$, one of these $2k$ instances causes high regret.

Let N_i be the number of bid-independent rounds in which agent i is played. Note that N_i does not depend on the bids. It is a random variable in the probability space induced by the clicks; its distribution is completely specified by the CTRs. We show that (in a certain sense) the allocation cannot distinguish between $\vec{\mu}_0$ and $\vec{\mu}_i$ if N_i is too small. Specifically, let $\mathcal{A}_t$ be the allocation in round t. Once the bids are fixed, this is a random variable in the probability space induced by the clicks. For a given set S of agents,

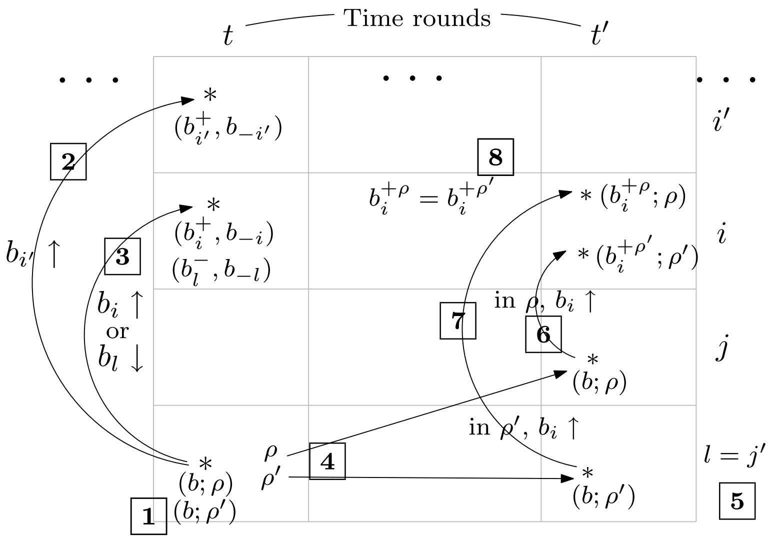

Figure 1: This figure explains all the steps in the proof of Lemma 3.9. The rows correspond to agents (whose identity is shown on the right side), and columns correspond to time rounds. The asterisks show the impressions. The arrows show how the impressions get *transferred*, and labels on the arrows show what causes the transfer. In labels, "in ρ, $b_i \uparrow$" denotes that a particular transfer of impression is caused in realization ρ when bid b_i in increased.

we consider the event $\{\mathcal{A}_t \in S\}$ for some fixed round t, and upper-bound the difference between the probability of this event under $\vec{\mu}_0$ and $\vec{\mu}_i$ in terms of $\mathbb{E}_i[N_i]$, in the following crucial claim, which is proved in the full version [8] via relative entropy techniques.

Claim 4.2. *For any fixed vector of bids, each round t, each agent i and each set of agents S, we have*

$$|\, \mathbb{P}_0[\mathcal{A}_t \in S] - \mathbb{P}_i[\mathcal{A}_t \in S] \,| \leq O(\epsilon^2 \, \mathbb{E}_0[N_i]). \tag{4.1}$$

Proof **(of Theorem 4.1).** Fix a positive constant β to be specified later. Consider the case $k = 2$ first. If $\mathbb{E}_0[N_i] > \beta T^{2/3}$ for some agent i, then on the problem instance $\mathcal{J}_i$, regret is $\Omega(T^{2/3})$. So without loss of generality let us assume $\mathbb{E}_0[N_i] \leq \beta T^{2/3}$ for each agent i. Then, plugging in the values for ϵ and $\mathbb{E}_0[N_i]$, the right-hand side of (4.1) is at most $O(\beta)$. Take β so that the right-hand side of (4.1) is at most $\frac{1}{4}$. For each round t there is an agent i such that $\mathbb{P}_0[\mathcal{A}_t \neq i] \geq \frac{1}{2}$. Then $\mathbb{P}_i[\mathcal{A}_t \neq i] \geq \frac{1}{4}$ by Claim 4.2, and therefore in this round algorithm $\mathcal{A}$ incurs regret $\Omega(\epsilon \, v_{\max})$ under problem instance $\mathcal{I}_i$. By Pigeonhole Principle there exists an i such that this happens for at least half of the rounds t, which gives the desired lower-bound.

Case $k \geq 3$ requires a different (and somewhat more complicated) argument. Let $R = \beta \, k^{1/3} \, T^{2/3}$ and N be the number of bid-independent rounds. Assume $\mathbb{E}_0[N] > R$. Then $\mathbb{E}_0[N_i] \leq \frac{1}{k} \, \mathbb{E}_0[N]$ for some agent i. For the problem instance $\mathcal{J}_i$ there are, in expectation, $E[N - N_i] = \Omega(R)$ bid-independent rounds in which agent i is not played; each of which contributes $\Omega(v_{\max})$ to regret, so the total regret is $\Omega(v_{\max} \, R)$.

From now on assume that $\mathbb{E}_0[N] \leq R$. Note that by Pigeonhole Principle, there are more than $\frac{k}{2}$ agents i such that $\mathbb{E}_0[N_i] \leq 2R/k$. Furthermore, let us say that an agent i is *good* if $\mathbb{P}_0[\mathcal{A}_t = i] \leq \frac{4}{5}$ for more than $T/6$ different rounds t. We claim that there are more than $\frac{k}{2}$ good agents. Suppose not. If agent i is not good then $\mathbb{P}_0[\mathcal{A}_t = i] > \frac{4}{5}$ for at least $\frac{5}{6} T$ different rounds t, so if there are at least $k/2$ such agents then

$$T = \textstyle\sum_{t=1}^{T} \sum_{i=1}^{k} \mathbb{P}_0[\mathcal{A}_t = i] > \frac{k}{2} \times (\frac{5}{6} T) \times \frac{4}{5} \geq kT/3 \geq T,$$

contradiction. Claim proved. It follows that there exists a good agent i such that $\mathbb{E}_0[N_i] \leq 2R/k$. Therefore the right-hand side of (4.1) is at most $O(\beta)$. Pick β so that the right-hand side of (4.1) is at most $\frac{1}{10}$. Then by Claim 4.2 for at least $T/6$ different rounds t we have $\mathbb{P}_i[\mathcal{A}_t = i] \leq \frac{9}{10}$. In each such round, if agent i is not played then algorithm $\mathcal{A}$ incurs regret $\Omega(\epsilon \, v_{\max})$ on problem instance $\mathcal{I}_i$. Therefore, the (total) regret of $\mathcal{A}$ on problem instance $\mathcal{I}_i$ is $\Omega(\epsilon \, v_{\max} \, T) = \Omega(v_{\max} \, k^{1/3} \, T^{2/3})$. □

Theorem 4.3. *In the setting of Theorem 4.1, fix k and $v_{\max}$ and assume that $R(T; v_{\max}) = O(v_{\max} \, T^{\gamma})$ for some $\gamma < 1$. Then for every fixed $\delta \leq \frac{1}{4}$ and $\lambda < 2(1 - \gamma)$ we have $R_{\delta}(T; v_{\max}) = \Omega(\delta \, v_{\max} \, T^{\lambda})$.*

Proof. Fix $\lambda \in (0, \, 2(1 - \gamma))$. Redefine $\vec{\mu}_i$'s with respect to a different ϵ, namely $\epsilon = T^{-\lambda/2}$. Define the problem instances $\mathcal{I}_i$ in the same way as before: all bids are $v_{\max}$, the CTRs are given by $\vec{\mu}_i$.

Let us focus on agents 1 and 2. We claim that $\mathbb{E}_1[N_1] + \mathbb{E}_2[N_2] \geq \beta \, T^{\lambda}$, where $\beta > 0$ is a constant to be defined later. Suppose not. Fix all bids to be $v_{\max}$. For each round t, consider event $S_t = \{\mathcal{A}_t = 1\}$. Then by Claim 4.2

$$\begin{aligned} \left| \mathbb{P}_1[S_t] - \mathbb{P}_2[S_t] \right| &\leq \left| \mathbb{P}_0[S_t] - \mathbb{P}_1[S_t] \right| + \left| \mathbb{P}_0[S_t] - \mathbb{P}_2[S_t] \right| \\ &\leq O\left(\epsilon^2\right) (\mathbb{E}_1[N_1] + \mathbb{E}_2[N_2]) \leq \tfrac{1}{4} \end{aligned}$$

for a sufficiently small β. Now, $\mathbb{P}_1[S_t] \geq \frac{1}{2}$ for at least $T/2$ rounds t. This is because otherwise on problem instance $\mathcal{I}_i$ regret would be $R(T) \geq \Omega(\epsilon \, T v_{\max}) = \Omega(v_{\max} \, T^{1-\lambda/2})$, which contradicts the assumption $R(T) = O(v_{\max} \, T^{\gamma})$. Therefore $\mathbb{P}_2[S_t] \geq \frac{1}{4}$ for at least $T/2$ rounds t, hence on problem instance $\mathcal{I}_2$ regret is at least $\Omega(\epsilon\, T v_{\max})$, contradiction. Claim proved.

Now without loss of generality let us assume that $\mathbb{E}_1[N_1] \geq \frac{\beta}{2}\, T^{\lambda}$. Consider the problem instance in which CTRs given by $\vec{\mu}_1$, bid of agent 2 is $v_{\max}$, and all other bids are $v_{\max}(1-2\delta)/(1+2\epsilon)$. It is easy to see that this problem instance has δ-gap. Each time agent 1 is selected, algorithm incurs regret $\Omega(\delta v_{\max})$. Thus the total regret is at least $\Omega(\delta N_1\, v_{\max}) = \Omega(\delta\, v_{\max}\, T^{\lambda})$. □

Matching upper bound. Let us describe a very simple mechanism, called *the naive MAB mechanism*, which matches the lower bound from Theorem 4.1 up to polylogarithmic factors (and also the lower bound from Theorem 4.3, for $\gamma = \lambda = \frac{2}{3}$ and constant δ). [9]

Fix the number of agents k, the time horizon T, and the bid vector b. The mechanism has two phases. In the *exploration phase*, each agent is played for $T_0 := k^{-2/3}\, T^{2/3} (\log T)^{1/3}$ rounds, in a round robin fashion. Let c_i be the number of clicks on agent i in the exploration phase. In the *exploitation phase*, an agent $i^* \in \operatorname{argmax}_i c_i b_i$ is chosen and played in all remaining rounds. Payments are defined as follows: agent i^* pays $\max_{i \in [k] \setminus \{i^*\}} c_i b_i / c_{i^*}$ for every click she gets in exploitation phase, and all others pay 0. (Exploration rounds are free for every agent.) This completes the description of the mechanism.

Observation 4.4. *Consider the stochastic MAB mechanism design problem with k agents. The naive mechanism is normalized, truthful and has worst-case regret $R(T; v_{\max}) = O(v_{\max}\, k^{1/3}\, T^{2/3}\, \log^{2/3} T)$.*

Proof. The mechanism is truthful by a simple second-price argument.[10] Recall that c_i is the number of clicks i got in the exploration phase. Let $p_i = \max_{j \neq i} c_j b_j / c_i$ be the price paid (per click) by agent i if she wins (all) rounds in exploitation phase. If $v_i \geq p_i$, then by bidding anything greater than p_i agent i gains $v_i - p_i$ utility each click irrespective of her bid, and bidding less than v_i, she gains 0, so bidding v_i is weakly dominant. Similarly, if $v_i < p_i$, then by bidding anything less than p_i she gains 0, while bidding $b_i > p_i$, she *loses* $b_i - p_i$ each click. So bidding v_i is weakly dominant in this case too.

For the regret bound, let $(\mu_1\,,\;\ldots\;, \mu_k)$ be the vector of CTRs, and let $\bar{\mu}_i = c_i / T_0$ be the sample CTRs. By Chernoff bounds, for each agent i we have $\Pr\left[|\bar{\mu}_i - \mu_i| > r\right] \leq T^{-4}$, for $r = \sqrt{8 \log(T)/T_0}$. If in a given run of the mechanism all estimates $\bar{\mu}_i$ lie in the intervals specified above, call the run *clean.* The expected regret from the runs that are not clean is at most $O(v_{\max})$, and can thus be ignored. From now on let us assume that the run is clean.

The regret in the exploration phase is at most $k\, T_0\, v_{\max} = O(v_{\max}\, k^{1/3}\, T^{2/3}\, \log^{1/3} T)$. For the exploitation phase, let $j = \operatorname{argmax}_i \mu_i b_i$. Then (since we assume that the run is clean) we have

$$(\mu_{i^*} + r)\, b_{i^*} \geq \bar{\mu}_{i^*}\, b_{i^*} \geq \bar{\mu}_j\, b_j \geq (\mu_j - r)\, b_j,$$

which implies $\mu_j v_j - \mu_{i^*} v_{i^*} \leq r(v_j + v_{i^*}) \leq 2r\, v_{\max}$. Therefore, the regret in exploitation phase is at most $2r\, v_{\max}\, T = O(v_{\max}\, k^{1/3}\, T^{2/3}\, \log^{2/3} T)$. Therefore the total regret is as claimed. □

5. EXTENSIONS AND OPEN QUESTIONS

We extend our results in several directions which are fleshed out in the full version [8].

First, we derive a regret lower bound for deterministic truthful mechanisms without assuming that the allocations are scale-free. In particular, for two agents there are no assumptions. This lower bound holds for any k (the number of agents) assuming IIA, but unlike the one in Theorem 4.1 it does not depend on k.[11]

Second, we extend our results to randomized mechanisms. We consider randomized mechanisms that are *universally truthful*, i.e. truthful for each realization of the internal random seed. For mechanisms that randomize over exploration-separated deterministic allocation rules, we obtain the same lower bounds as in Theorems 4.1 and Theorem 4.3.

Third, we consider randomized allocation rules under a weaker version of truthfulness: a mechanism is *weakly truthful* if for each realization, it is truthful in expectation over its random seed. We show that any randomized allocation that is "pointwise monotone" and satisfies a certain notion of "separation between exploration and exploitation" can be turned into a mechanism that is weakly truthful and normalized. Then we apply this result to two algorithms in the literature [22, 14] in order to obtain regret guarantees for the version of the MAB mechanism design problem in which the clicks are chosen by an adversary. (This version corresponds to the *adversarial MAB problem* [7, 14, 1, 9].) In particular, for oblivious (resp. adaptive) adversaries the upper bound matches our lower bound for deterministic allocations up to $(\log k)^{1/3}$ (resp. $k^{2/3}$) factors.

Fourth, we consider the stochastic MAB mechanism design problem under a more relaxed notion of truthfulness: truthfulness *in expectation*, where for each vector of CTRs the expectation is taken over clicks (and the internal randomness in the mechanism, if the latter is not deterministic). Following our line of investigation, we ask whether restricting a mechanism to be truthful in expectation has any implications on the structure and regret thereof. Given our results on mechanisms that are truthful and normalized, it is tempting to seek similar results for mechanisms that are truthful in expectation and normalized in expectation.[12] We rule out this approach: we show that in order to obtain any non-trivial lower bounds on regret and (essentially) any non-trivial structural results, one needs to assume that a mechanism is ex-post normalized, at least in some ap-

[9] Independently, Devanur and Kakade [15] presented a version of the naive MAB mechanism that achieves the same regret even in the more general model in which the value-per-click of an agent changes over time and the agents are allowed to submit a different bid at every round. Instead of assigning all impressions to the same agent in the exploitation phase, their mechanism runs the same allocation and payment procedure for each exploration round separately (see [15] for details).

[10] Alternatively, one can use Theorem 3.8 since all exploration rounds are bid-independent, and only exploration rounds are influential, and the payments are exactly as defined in Theorem 3.1.

[11] One would expect to obtain such bound by a reduction to the two-agent case. Interestingly, the trivial reduction fails.

[12] A mechanism is *normalized in expectation* if in expectation over clicks (and possibly over the allocation's randomness), each agent is charged an amount between 0 and her bid for each click she receives.

proximate sense. The key idea is to view the allocation and the payment as multivariate polynomials over the CTRs.

The two major questions left open by this work concern structural results and regret lower bounds for (i) weakly truthful randomized mechanisms allocations, and (ii) mechanisms that are truthful in expectation. The latter question seems to require very different techniques which would further explore the connection to polynomials over the CTRs. Another potentially fruitful line of inquiry concerns incorporating more detailed settings, such as: budget constraints, time-varying valuations, repeated bids, and external partial information on CTRs.

Acknowledgements. We thank Jason Hartline, Robert Kleinberg and Ilya Segal for helpful discussions.

6. REFERENCES

[1] Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Conf. on Learning Theory (COLT)*, pages 263–274, 2008.

[2] Gagan Aggarwal, Ashish Goel, and Rajeev Motwani. Truthful auctions for pricing search keywords. In *ACM Conf. on Electronic Commerce (EC)*, pages 1–7, 2006.

[3] Gagan Aggarwal and S. Muthukrishnan. Tutorial on theory of sponsored search auctions. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2008.

[4] Aaron Archer and Éva Tardos. Truthful mechanisms for one-parameter agents. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 482–491, 2001.

[5] Susan Athey and Ilya Segal. An efficient dynamic mechanism. Available from `http://www.stanford.edu/~isegal/agv.pdf`, March 2007.

[6] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2–3):235–256, 2002. Preliminary version in *15th ICML*, 1998.

[7] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77, 2002. Preliminary version in *36th IEEE FOCS*, 1995.

[8] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. Characterizing Truthful Multi-Armed Bandit Mechanisms . Technical Report `http://arxiv.org/abs/0812.2291`, December 2008. Minor revision February 2009.

[9] Peter L. Bartlett, Varsha Dani, Thomas Hayes, Sham Kakade, Alexander Rakhlin, and Ambuj Tewari. High-probability regret bounds for bandit online linear optimization. In *Conf. on Learning Theory (COLT)*, pages 335–342, 2008.

[10] Michael Ben-Or and Avinatan Hassidim. The Bayesian Learner is Optimal for Noisy Binary Search (and Pretty Good for Quantum as Well). In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2008.

[11] Dirk Bergemann and Juuso Välimäki. Efficient dynamic auctions. Available from `cowles.econ.yale.edu/P/cd/d15b/d1584.pdf`, October 2006.

[12] Donald Berry and Bert Fristedt. *Bandit problems: sequential allocation of experiments.* Chapman&Hall, 1985.

[13] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games.* Cambridge University Press, 2006.

[14] Varsha Dani and Thomas P. Hayes. Robbing the bandit: Less regret in online geometric optimization against an adaptive adversary. In *16th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 937–943, 2006.

[15] Nikhil Devanur and Sham M. Kakade. The price of truthfulness for pay-per-click auctions. In *10th ACM Conf. on Electronic Commerce (EC)*, 2009.

[16] Shahar Dobzinski and Mukund Sundararajan. On characterizations of truthful mechanisms for combinatorial auctions and scheduling. In *ACM Conf. on Electronic Commerce (EC)*, pages 38–47, 2008.

[17] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97(1):242–259, March 2007.

[18] J. C. Gittins. *Multi-Armed Bandit Allocation Indices.* John Wiley & Sons, 1989.

[19] Rica Gonen and Elan Pavlov. An incentive-compatible multi-armed bandit mechanism. In *ACM Symp. on Principles Of Distributed Computing (PODC) (Brief Announcement)*, pages 362–363, 2007. Preliminary version in *3rd Workshop on Sponsored Search Auctions* (in conjunction with *WWW* 2007).

[20] Nicole Immorlica, Kamal Jain, Mohammad Mahdian, and Kunal Talwar. Click fraud resistant methods for learning click-through rates. In *Intl. Workshop On Internet And Network Economics (WINE)*, pages 34–45, 2005.

[21] Richard Karp and Robert Kleinberg. Noisy binary search and its applications. In *18th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 881–890, 2007.

[22] Robert Kleinberg. Lecture notes: *CS683: Learning, Games, and Electronic Markets* (week 8), Spring 2007. Available at http://www.cs.cornell.edu/courses/cs683/2007sp.

[23] Robert Kleinberg. Lecture notes: *CS683: Learning, Games, and Electronic Markets* (week 9), Spring 2007. Available at http://www.cs.cornell.edu/courses/cs683/2007sp.

[24] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-Armed Bandits in Metric Spaces. In *40th ACM Symp. on Theory of Computing (STOC)*, pages 681–690, 2008.

[25] Sebastien Lahaie, David M. Pennock, Amin Saberi, and Rakesh V. Vohra. *In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani (eds.) Chapter 28, Sponsored search auctions.* Cambridge University Press., 2007.

[26] T.L. Lai and Herbert Robbins. Asymptotically efficient Adaptive Allocation Rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

[27] Ron Lavi, Ahuva Mu'alem, and Noam Nisan. Towards a characterization of truthful combinatorial auctions. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, page 574, 2003.

[28] Ron Lavi and Noam Nisan. Online ascending auctions for gradually expiring items. In *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 1146–1155, 2005.

[29] Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. Adwords and generalized online matching. *J. ACM*, 54(5):22, 2007.

[30] Roger B. Myerson. Optimal Auction Design. *Mathematics of Operations Research*, 6:58–73, 1981.

[31] Hamid Nazerzadeh, Amin Saberi, and Rakesh Vohra. Dynamic cost-per-action mechanisms and applications to online advertising. In *17th Intl. World Wide Web Conf. (WWW)*, pages 179–188, 2008.

[32] N. Nisan and A. Ronen. Algorithmic Mechanism Design. *Games and Economic Behavior*, 35(1-2):166–196, 2001.

[33] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani (eds.). *Algorithmic Game Theory.* Cambridge University Press., 2007.

[34] Christos Papadimitriou, Michael Schapira, and Yaron Singer. On the hardness of being truthful. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2008.

[35] Tim Roughgarden. An algorithmic game theory primer. IFIP International Conference on Theoretical Computer Science (TCS). An invited survey., 2008.

[36] Hal R. Varian. Position auctions. *International Journal of Industrial Organization*, 25(6):1163–1178, December 2007.