Truthful multi-armed bandit mechanisms for multi-slot sponsored search auctions

Akash Das Sarma[1], Sujit Gujar[2] and Y. Narahari[3,*]

[1]Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur 208 016, India
[2]Xerox Research Centre India, Bangalore 560 066, India
[3]Department of Computer Science and Automation, Indian Institute of Science, Bangalore 560 012, India

In pay-per-click sponsored search auctions which are currently extensively used by search engines, the auction for a keyword involves a certain number of advertisers (say *k*) competing for available slots (say *m*) to display their advertisements (ads for short). A sponsored search auction for a keyword is typically conducted for a number of rounds (say *T*). There are click probabilities μ_{ij} associated with each agent–slot pair (agent *i* and slot *j*). The search engine would like to maximize the social welfare of the advertisers, that is, the sum of values of the advertisers for the keyword. However, the search engine does not know the true values advertisers have for a click to their respective advertisements and also does not know the click probabilities. A key problem for the search engine therefore is to learn these click probabilities during the initial rounds of the auction and also to ensure that the auction mechanism is truthful. Mechanisms for addressing such learning and incentives issues have recently been introduced. These mechanisms, due to their connection to the multi-armed bandit problem, are aptly referred to as multi-armed bandit (MAB) mechanisms. When $m = 1$, exact characterizations for truthful MAB mechanisms are available in the literature. Recent work has focused on the more realistic but non-trivial general case when $m > 1$ and a few promising results have started appearing. In this article, we consider this general case when $m > 1$ and prove several interesting results. Our contributions include: (1) When μ_{ij}s are *unconstrained*, we prove that any truthful mechanism must satisfy *strong pointwise monotonicity* and show that the regret will be $\Theta(T)$ for such mechanisms. (2) When the clicks on the ads follow a certain click *precedence property*, we show that *weak pointwise monotonicity* is necessary for MAB mechanisms to be truthful. (3) If the search engine has a certain coarse pre-estimate of μ_{ij} values and wishes to update them during the course of the *T* rounds, we show that *weak pointwise monotonicity* and *type-I separatedness* are necessary while *weak pointwise monotonicity* and *type-II separatedness* are sufficient conditions for the MAB mechanisms to be truthful. (4) If the click probabilities are separable into agent-specific and slot-specific terms, we provide a characterization of MAB mechanisms that are *truthful in expectation*.

Keywords: Bandit mechanisms, click probability, search auctions, search engines.

Introduction

IN the last two decades, use of the Internet has increased exponentially and Internet advertising has become the most effective way of advertisements due to its immediate reach to billions of people. The advertisements (referred to as *ads* for brevity) of various businesses are displayed at different locations on the websites. For more about Internet advertising, see refs 1–4.

Sponsored search auctions

Popular search engines such as Google, Yahoo! and Bing offer search facility for free. Whenever a user searches for any set of keywords on a search engine, along with the search results, called as *organic results*, the search engine displays advertisements related to those keywords on the right side of the organic results or at the top of the organic results (Figure 1). Such ads carry a hyper-link called *sponsored link* to the advertiser's website. The search engine charges the advertiser for displaying her ad. One model of payment is the *pay-per-impression* model in which the search engine charges the advertiser each time it displays her ad. The more prevalent model is *pay-per-click*. In this model, the search engine charges an advertiser for displaying her ad only if a user clicks on her ad.

Typically, there are more advertisers than slots available for displaying ads. Considering the competition for the sponsored links, the search engine usually employs an auction mechanism. The decision regarding which ads are to be displayed and their respective order is based on the bids submitted by the advertisers indicating the maximum amount they are willing to pay per click. So for each keyword, there is an auction that takes place in the background. These auctions are called *sponsored search*

*For correspondence. (e-mail: hari@csa.iisc.ernet.in)

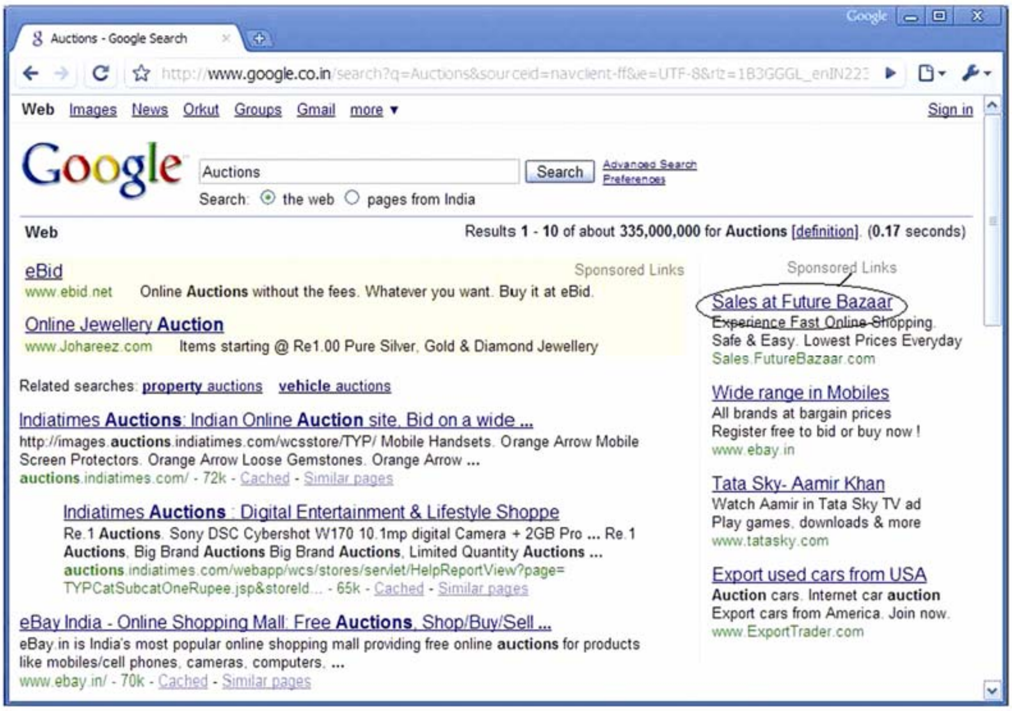

Figure 1. Sponsored slots in a search page.

auctions or pay-per-click auctions. Several varieties of auction mechanisms have been proposed for sponsored search auctions. Each auction mechanism has an *allocation rule* and a *payment rule*. The allocation rule determines which advertisements will be chosen to appear in the available slots and the actual allocation of slots to advertisers. The payment rule determines the amount that an allocated advertiser will pay each time the corresponding slot is clicked by a user.

The auction mechanism that was first used in this context was the generalized first price (GFP) mechanism. The advertising slots are allocated to the advertisers in *descending order of their bids*. If two advertisers place the same bid, then the tie is broken by an appropriate rule. Every time a user clicks on a sponsored link, an advertiser's account is automatically billed the amount of the advertiser's bid.

Most search engines currently use a mechanism that is based on the generalized second price (GSP) mechanism. In the GSP auction mechanism, the allocation rule is the same as in GFP. However, every time a user clicks on a sponsored link, an advertiser's account is automatically billed the amount of the advertiser's bid who is just below the advertiser in the ranking of the displayed advertisements plus a minimum increment. Many other mechanisms have also been discussed in the literature. For more details on the sponsored search auctions, the reader is referred to Narahari *et al.*[5].

To perform any optimizations, such as maximizing social welfare of the advertisers or maximizing revenue to the search engine, the true valuations of the advertisers are needed. Being rational, the advertisers may actually manipulate their bids and therefore a primary goal of the search engine is to design an auction for which it is in the best interest of each advertiser to bid truthfully irrespective of the bids of the other advertisers. Such an auction is said to be *dominant strategy incentive compatible* (DSIC), or truthful.

Multi-armed bandit mechanisms

A click on a displayed ad by a user is a random event. The probability of the ad getting clicked depends upon the advertiser as well as in what position the ad is displayed. These click probabilities or clickthrough rates (CTRs), play a crucial role in sponsored search auctions. Given an agent i and a slot j, let us define the click probability μ_{ij} as the probability with which the ad of an agent i will be clicked if the ad appears in slot j. If the search engine knows the CTRs, then its problem is only to design a truthful auction. However, the search engine may not know the CTRs beforehand. Thus the problem of the search engine is twofold: (1) learn the CTR values and (2) design a truthful auction.

Typically, the same set of agents would be involved for a given keyword. The search engine can exploit this fact to learn the CTRs by initially displaying ads by the advertisers involved. Also it is reasonable to assume that the advertisers may not revise their bids frequently. If the advertisers were bidding true values, the search engine's problem would be the same as that of a multi-armed bandit (MAB) problem for learning the CTRs. A brief description of the multi-armed bandit problem follows.

MAB problem: In this problem, there is a slot machine with k different arms. When an arm is pulled, a probabilistic reward accrues. The problem is to decide which of the k arms is to be pulled in multiple plays of the game. The decision-maker pulls the arms multiple times and has to optimize his total reward based on gained information as well as gained knowledge about the available rewards. The MAB problem was first studied by Robbins[6] in 1952. If the expected rewards are known, then the obvious choice would be to pull the arm with the highest expected reward. However, the player has to learn the expected rewards while pulling the different arms in different rounds. The decision-maker's dilemma is to choose the arm which is expected to yield the highest rewards from experience or try out some other arms which may give better rewards. The first strategy is known as *exploitation* from the previous knowledge and the second strategy is known as *exploration*.

For the infinite horizon case, an index policy was introduced by Gittins *et al.*[7], which is known as the *Gittins index*. There have been several generalizations of the Gittins index policies. An algorithm or policy is a recommendation of which arm is to be pulled in each round. The Gittins index-based policy assumes that the decision-maker can compute the payoffs he/she would receive from each of the arms ahead of time[8]. Thus, the problem is more of optimization than exploration–exploitation[8]. The Gittins index-based policy can be generalized with suitable assumptions on distribution of rewards. The algorithm of Lai and Robbins[9] works with assumptions about the stationary distribution on rewards. The most general analysis and algorithms for the finite horizon case can be found in Auer *et al.*[8]. The performance of the MAB algorithms is measured in terms of regret, that is, the difference between reward that could have been achieved by pulling the best arm in each round minus the reward actually achieved by pulling arms according to the given algorithm. Auer *et al.*[8] showed that even the best possible algorithm with no statistical assumptions on rewards can achieve a regret no less than $O(\sqrt{T \log T})$, where T is the number of rounds.

The MAB problems have been quite useful in numerous applications in economics and engineering, especially, in market learning, pricing, multi-agent experimentation, etc. A survey of algorithms for solving different versions of the MAB problem is presented in ref. 10.

MAB mechanisms: In the above setting of the MAB problem, it is assumed that the reward produced when an arm is pulled, though random, is observable by the decision-maker. If the reward produced is determined by a strategic agent (such as an advertiser in sponsored search auction), the reward is private information of the agent and may not be reported truthfully by the agent. This leads to a mechanism design problem where the decision-maker not only has to extract true rewards (or values that each worker assigns for each click) from the agents, but also has to learn the actual clickthrough rates. Such mechanisms in the context of the MAB problem are called MAB mechanisms. They are very apt for modelling the problem facing the search engine in sponsored search auctions. The problem of the search engine is one of designing an incentive-compatible MAB mechanism. In the initial rounds, the search engine displays advertisements from all the agents to learn the CTRs (*exploration phase*). Then it uses the information gained in these rounds to maximize the social welfare (*exploitation phase*), where social welfare is the sum of valuations of the advertisers who are allocated slots. The search engine will invariably lose a part of social welfare for the exploration phase. The difference between the social welfare the search engine would have achieved with the knowledge of CTRs and the actual social welfare achieved by a MAB mechanism is referred to as regret. Thus, *regret* analysis is also important while designing an MAB mechanism.

Related work

After the seminal work of Robbins[6], MAB problems have been extensively studied for regret analysis and convergence rates. Readers are referred to Auer *et al.*[8] for regret analysis in finite time MAB problems. However, when a mechanism designer has to consider strategic behaviour of the agents, these bounds on regret would not apply anymore. Babaioff *et al.*[11], have derived a characterization for truthful MAB mechanisms in the context of pay-per-click sponsored search auctions if there is only a single slot for each keyword. They have shown that any truthful MAB mechanism must have at least $\Omega(T^{2/3})$ worst case regret and also proposed a mechanism that achieves this regret. Here T indicates the number of rounds for which the auction is conducted for a given keyword, with the same set of agents involved.

Devanur and Kakade[12] have also addressed the problem of designing truthful MAB mechanisms for pay-per-click auctions with a single sponsored slot. Though they have not explicitly attempted a characterization of truthful MAB mechanisms, they have derived similar results on payments as in Babaioff *et al.*[11]. They have also obtained a bound on regret of a MAB mechanism to be $\Theta(T^{2/3})$. Note that the regret in Devanur and Kakade[12] is regret in the revenue to the search engine, as against regret analysis in Babaioff *et al.*[11] which is for social welfare of the advertisers. In this article, unless explicitly stated, when we refer to *regret*, we mean loss in the social welfare as compared to the social welfare that could have been obtained with the known CTRs.

Babaioff *et al.*[13] showed that, if we use randomized mechanisms as against deterministic in the above papers, we can achieve regret that is, $O(T^{1/2})$ which is the best possible bound for randomized MAB algorithms too.

The notion of truthfulness used in the above papers is dominant strategy incentive compatibility. Zoeter[14] considered the learning of CTRs when agents can reincarnate themselves and considered the weaker notion of incentive compatibility, namely Bayesian incentive compatibility, to optimally learn the CTRs and avoid reincarnation by the advertisers with a new identity. This paper[14] does not perform any regret analysis.

In all the above papers, though the results are important, only a single slot for advertisements is considered. In real-world sponsored search auctions, typically there are multiple slots available for displaying the ads. Generalization of their work to the more realistic case of multiple sponsored slots is non-trivial and our article seeks to fill this research gap.

Gonen and Pavlov[15] addressed the issue of unknown CTRs in multiple slot sponsored search auctions and proposed a specific mechanism. Their claim that their mechanism is truthful in expectation has been contested[11,12]. Also Gonen and Pavlov[15] do not provide any characterization for truthful multi-slot MAB mechanisms.

More recently, Gatti *et al.*[16] have generalized the work of Devanur and Kakade[12] to the case of multi-slot sponsored search auctions. They obtained analytically upper bounds in revenue loss with respect to a Vickrey–Clarke–Groves auction and presented simulation results that investigated the accuracy of the upper bounds on predicting the dependency of the regret on the number of rounds, number of slots and number of advertisements. They proposed a modification to VCG auction in the setting of separable CTRs, that is the setting where the clickthrough rate is composed of two parts, one agent-dependent and the other slot-dependent (we will be discussing separable CTRs in more detail in a later section). The authors proved that their mechanism has regret $O(T^{2/3})$ which confirms the empirical evidence provided in our article. It is to be noted that the work of Gatti *et al.*[16] does not characterize the mechanisms that can be implemented truthfully.

Our contributions

In this article, we extend the results of Babaioff *et al.*[11], to the general case of two or more sponsored slots. The precise question we address is: *which MAB mechanisms for multi-slot pay-per-click sponsored search auctions are dominant strategy incentive compatible?* We describe our specific contributions below.

In the first and most general setting, we assume no knowledge of clickthrough rate (μ_{ij}) values or any relationships among μ_{ij} values. We refer to this setting as the 'unknown and unconstrained CTR' setting. Here we show that any truthful mechanism must satisfy a highly restrictive property which we refer to as *strong pointwise monotonicity* property. We show that all mechanisms satisfying this property will however exhibit a high regret, which is $\Theta(T)$. Then we explore the following variants of the general setting to obtain more specific characterizations.

First, we consider a setting where the realization is restricted according to a property that we call the *Higher slot click precedence* property (a click in a lower slot will automatically imply that a click is received if the same ad is shown in any higher slot). For this setting, we provide a weaker necessary condition than strong pointwise monotonicity. Finding a necessary and sufficient condition however remains open.

Next we provide a complete characterization of MAB mechanisms which are *truthful in expectation* under a stochastic setting where a coarse estimate of μ_{ij} is known to the auctioneer and to the agent *i*, perhaps from some database of past auctions. Under this setting, the auctioneer updates his database of μ_{ij} values based on the observed clicks, thereby improving his estimate and maximizing revenue.

Finally, we derive a complete characterization of truthful multi-slot MAB mechanisms for a stochastic setting where we assume that the μ_{ij}s are separable into agent-dependent and slot-dependent parts. Here, unlike the previous setting, we do not assume existence of any information on agent-dependent click probabilities.

For all the above multi-slot sponsored search auction settings, we show that the slot allocation in truthful mechanisms must satisfy some notion of monotonicity with respect to the agents' bids and a certain weak separation between exploration and exploitation. Our results are summarized in Table 1.

Our approach and line of attack in this article follow that of Babaioff *et al.*[11] where the authors used the notions of pointwise monotonicity, *weakly separatedness, and exploration separatedness* quite critically in characterizing truthfulness. Since our article deals with the general problem of which theirs is a special case, these notions continue to play an important role in our article. However, there are some notable differences as explained below. We generalize their notion of *pointwise monotonicity* in two ways. The first notion we refer to as strong pointwise monotonicity and the second one as weak pointwise monotonicity. In addition, we introduce the key notions of *influential set*, *i-influentiality* and *strongly influential*. We use these new notions to define a non-trivial generalization of their notion of *weakly separatedness*, into two types of separatedness, type-I and type-II separated. The characterization of truthful mechanisms for a single parameter was provided by Archer and Tardos[17] and Myerson[18]. For deriving payments to be

Table 1. Multi-armed bandit results

Number of slots (m)	Nature of the learning parameter CTR	Solution concept	Allocation rule	Regret
$m = 1$ (ref. 11)	Unrestricted	DSIC	Pointwise monotone and exploration separated	$\Theta(T^{2/3})$
$m > 1$	Unrestricted	DSIC	Strongly pointwise monotone and type-I separated	$\Theta(T)$
	Higher slot click precedence	DSIC	Weakly pointwise monotone and type-I separated (necessary condition)	Regret analysis not carried out
	CTR pre-estimates available	Truthful in expectation	Weakly pointwise monotone and type-I separated (necessary) and type-II separated (sufficient)	Regret analysis not carried out
	Separable CTR	Truthful in expectation	Weakly pointwise monotone and type-I separated (necessary) type-II separated (sufficient)	$\Omega(T^{2/3})$ (experimental evidence)

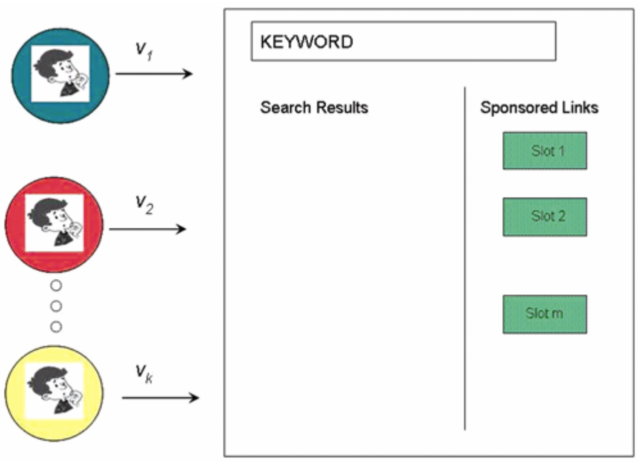

Figure 2. A sponsored search auction scenario.

assigned to the agents for truthful implementation, we use the approach of these authors[17,18].

In a later section we provide some experimental results on regret analysis, followed by conclusion.

System setup and notation

In the auction considered, there are k agents and m ad slots ($k \geq m$). Each agent has a single advertisement that she wants to display and a private value v_i which is her value per click on the ad. The auctioneer, that is the search engine, wishes to distribute the ads among these slots. This scenario is captured in Figure 2. The advertisements have certain click probabilities which depend upon the agent as well as the slot with which the agent is associated. Let μ_{ij} be the probability of an ad of an agent i receiving click in slot j. Now, the goal of the search engine is to assign these agents to the slots in a way that the social welfare, which is the total value received by the bidders, is maximized. However, there are two problems: (i) the search engine does not know v_i, the valuations of the agents and (ii) the search engine may not know the click probabilities μ_{ij}.

So, the goal of the search engine is: (i) to design a DSIC auction in which it is in the agents' interest to bid their true values, v_is (ii) to estimate μ_{ij}. We consider multi-round auctions, where the search engine displays the various advertisements repeatedly over a large number of rounds. The mechanism uses the initial rounds in an *explorative* fashion to learn μ_{ij} and then uses the other rounds *exploitatively* to gain value.

The system works as follows. At the start of the auction, each agent submits a sealed bid b_i. Based on this bid and the click information from previous rounds, the mechanism decides to allocate each ad slot to a particular agent and then displays the m chosen ads. The users can now click on any number of these ads and this information gets registered by the mechanism for future rounds. At the end of T rounds, depending on the bids submitted by the agents and the number of clicks received by each agent, the agents have to make a certain payment P_i to the mechanism.

Note: *P_i and C_i are functions of b and ρ. Whenever the arguments are clear from the context, we just refer to them as P_i and C_i.*

A mechanism can be formally defined as the tuple (A, P), where A is the allocation rule specifying the slot allocation and P is the payment rule.

The important notation used in the article is summarized in Table 2. We now define the terms used in this article.

We assume $T > 1$ to avoid corner cases in the proofs.

Table 2. Notation: multi-armed bandit mechanisms

K	$= \{1, 2, \ldots, k\}$, set of agents
M	$= \{1, 2, \ldots, m\}$, set of slots
i	Index of an agent, $i = 1, 2, \ldots, k$
j	Index of a slot, $j = 1, 2, \ldots, m$
T	Total number of rounds
t	A particular round $t \in \{1, 2, \ldots, T\}$
$\rho_{ij}(t)$	= 1 if agent i gets a click in slot j in round t, = 0 otherwise
$\rho(t)$	$(\rho_{ij}(t))_{i\in K, j\in M}$
ρ	$= (\rho(1), \rho(2), \ldots, \rho(T))$
v_i	Agent i's valuation of a click to her ad
b_i	Bid by agent i
b_{-i}	Bid vector of bids of all the agents, except i
b	Bid vector, indicating bids of all the agents $= (b_i, b_{-i}) = (b_1, b_2, \ldots, b_k)$
$A_{ij}(b, \rho, t)$	= 1 if an agent i is allocated slot j in round t, = 0 otherwise
$A(b, \rho, t)$	$(A_{ij}(b, \rho, t))_{i\in K, j\in M}$
$A(b, \rho)$	$= (A(b, \rho, 1), A(b, \rho, 2), \ldots, A(b, \rho, T))$, allocation rule
$C_i(b, \rho)$	Total number of clicks obtained by an agent i in T rounds
$P_i(b, \rho)$	Payment made by agent i
$P(b, \rho)$	$= (P_1(\cdot), P_2(\cdot), \ldots, P_k(\cdot))$, payment rule
$U_i(v_i, b, \rho)$	Utility of an agent i in T rounds $= v_iC_i(b, \rho) - P_i(b, \rho)$
b_i^+	A real number $> b_i$
α_i	Click probability associated with agent i
β_j	Click probability associated with slot j
μ_{ij}	The probability that an ad of an agent i receives click when the agent is allotted slot j
$N(b, \rho, i, t)$	Set of slot–agent pairs in round t that influence agent i in some future rounds
CTR	Clickthrough rate (click probability)
DSIC	Dominant strategy incentive compatible

Important notions and definitions

Definition 1. (*realization* ρ). *We define a realization* ρ *as a vector* $(\rho(1), \rho(2), \ldots, \rho(T))$, *where* $\rho(t) = [\rho_{ij}(t)]_{K\times M}$ *is click information in round t.* $\rho_{ij}(t) = 1$, *if an agent i's ad receives a click in slot j in round t, else* 0.

It is to be noted that the mechanism observes only those $\rho_{ij}(t)$ where $A_{ij}(b, \rho, t) = 1$. Thus, ρ is never completely known to the auctioneer even after execution of all T rounds. Also, though we use $A_{ij}(b, \rho, t)$ to indicate whether agent i is allocated in slot j in round t, it cannot depend upon future realizations. To be more precise, it should be interpreted as $A_{ij}(b, \rho(1), \rho(2), \ldots, \rho(t-1), t)$.

Definition 2. (*clickwise monotonicity*). *We call an allocation rule A clickwise monotone if for each agent i, for a fixed* (b_{-i}, ρ), *the number of clicks*, $C_i(b_i, b_{-i}, \rho)$ *is a non-decreasing function of* b_i. *That is*, $(\mathrm{d}C_i(.)/\mathrm{d}b_i) \geq 0$, $\forall (b_{-i}, \rho)$.

Definition 3. (*weak pointwise monotonicity*). *We call an allocation rule weak pointwise monotone if, for each agent i, for any given* (b_{-i}, ρ), *and bid* $b_i^+ > b_i$, $A_{ij}((b_i, b_{-i}), \rho, t) = 1 \Rightarrow A_{ij'}((b_i^+, b_{-i}), \rho, t) = 1$ *for some slot* $j' \leq j$, $\forall t$.

Example. *Consider a setting with four agents and two slots for advertisements for a particular keyword and T = 1,000. Let the click probabilities, CTRs, for each agent be decreasing. That is, for each agent, the lower slot has lower CTR. An allocation A be as follows. For first 100 rounds the advertisements of the four agents are displayed in each slot in a round-robin fashion. That is, in the first round, the ads of the agents 1 and 2 are shown in the slots 1 and 2 respectively, followed by in the slots 2 and 1 in next round. Then the agents 3 and 4 for next two rounds and again 1 and 2 till 100 rounds. For the remaining 900 rounds, the advertisements are displayed that maximize the expected sum of valuations of the clicks, where expectation is taken over the estimated click probabilities. If any of the bidders increases her bid, she will be displayed in higher slot. Thus, the allocation rule A described here is weakly pointwise monotone.* □

Definition 4. (*influential set*). *Given a bid vector*, b, *a realization* ρ *and round t, an influential set* $I(b, \rho, t)$ *is the set of all agent–slot allocation pairs* (i, j), *such that* (i) $A_{ij}(b, \rho, t) = 1$ *and* (ii) *a change in* $\rho_{ij}(t)$ *will result in a change in the allocation in some future round. t is referred to as an influential round. Agent i is referred to as an influential agent and j as influential slot w.r.t round t.*

As the mechanism learns CTRs while executing repeated auctions, it updates CTRs based on the observed clicks. As the mechanism updates CTRs, some of the allocations may change. The observed clicks which change some future allocation are said to influence some future allocation. Agent–slot pair (i, j) is influential in round t, if $A_{ij}(b, \rho, t) = 1$ and in some future round t', the allocation is changed based on $\rho_{ij}(t) = 0$ or 1 if other observed clicks are same till t'.

Definition 5. (*i-influential set*). *We define the i-influential set $N(b, \rho, i, t) \subseteq I(b, \rho, t)$ as the set of all influential agent–slot pairs (i', j') such that change in $\rho_{i'j'}(t)$ will change the allocation of agent i in some future round.*

Definition 6. (*strongly influential*). *We call a slot–agent pair (i^*, j^*) strongly influential in round t w.r.t. the realization $\rho(t)$, if changing the realization (toggling) in the bit $\rho_{i^*j^*}(t)$ changes the allocation in a future round. We call such a set (i^*, j^*, t) strongly i-influential if one of its influenced agents is i.*

Definition 7. (*type-I separated*). *We call an allocation rule type-I separated if for a given (b_{-i}, ρ), if $N((b_i, b_{-i}), \rho, i, t)$ is i-influential set, then $\forall(i', j') \in N((b_i, b_{-i}), \rho, i, t)$, $A_{i'j'} = 1$ when the agent i increases her bid to b_i^+.*

This means that when an agent i increases her bid, while the other parameters are kept fixed, the allocation in the originally influential slots does not change.

Definition 8. (*type-II separated*). *We call an allocation rule type-II separated if for a given (b_{-i}, ρ) and two bids of agent i, b_i and $b_i^+ > b_i$, $N((b_i, b_{-i}), \rho, i, t) \subseteq N((b_i^+, b_{-i}), \rho, i, t)$.*

This means that when an agent i increases her bid, while the other parameters are kept fixed, the allocation in the originally influential slots does not change and they remain influential. We will explain the above definitions with an example.

Example. (*Weakly separated allocation rule*). *Suppose there are four agents competing for displaying their ads in any of the two slots available for two rounds. An allocation rule A is defined as, in the first round, the ad of the agent i is displayed in the slot i, $i = 1, 2$. If any of the ads receives a click, that ad is retained in round the 2. If the ad in slot i is not clicked and if $b_i < b_{i+2}$, then in round 2, the ad of the agent $i + 2$ is displayed in slot i, else the original ad is retained. Now, assume, $b_1 < b_3$ and $b_2 < b_4$. Thus, if agent 1 receives the click in round 1, then she retains the slot in the second round, else she loses the slot. Thus, she is influencing herself in round 2. Similarly, agent 2 is influential for herself. Thus, when $b_1 < b_3$ and $b_2 < b_4$, $I(b, 1) = \{(1, 1), (2, 2)\}$. $N(b, 1, 1) = \{(1, 1)\}$ and $N(b, 2, 1) = \{(2, 2)\}$. If agent 1 or 2 increases her bid, still in round 1, she retains her slot. So, A is type-I separated. However, if $b_1 > b_3$, then she is not influential for herself. Thus, A is not type-II separated.* □

Note, type-II separated rule is also a type-I separated rule. We show that any truthful allocation rule must be type-I separated. However, for sufficiency, we use type-II separatedness property of an allocation rule.

We continue to use definitions of *normalized mechanism* and *non-degeneracy* from Babaioff *et al.*[11].

Definition 9. (*non-degeneracy*[11]). *An allocation rule is said to be non-degenerate if for any given realization ρ and bid profile (b_i, b_{-i}) there exists a finite interval X around b_i such that the allocation in all rounds is the same for any bid profile (x, b_{-i}) where $x \in X$.*

Definition 10. (*normalized mechanism*[11]). *A mechanism is said to be normalized if the payment rule is defined such that each agent i pays at most b_i for each click that she gets.*

Note, the definitions 3–9 are the properties of an allocation rule, while definition 10 is a property of a mechanism.

With these preliminaries, we are now ready to characterize truthful MAB mechanisms for various settings in the next section.

Characterization of truthful MAB mechanisms

Before stating our results, we prove a minor claim that we will use to develop our characterizations. We will use this claim implicitly in our proofs.

Claim 1. *Given $(b, (\rho(1), \rho(2), \ldots, \rho(t-1)))$, if (i^*, j^*) is i-influential in round t, then $\exists \rho^*(t)$ such that (i^*, j^*) is also strongly i-influential w.r.t. $\rho^*(t)$ in round t.*

Proof. Suppose the claim is false. Let the i-influential set of slots in round t be

$$N(b, \rho, i, t) = \{(i^1, j^1), (i^2, j^2), \ldots, (i^l, j^l), (i^*, j^*)\}.$$

It can be seen that $N(b, \rho, i, t) \neq \phi$ since it has at least one element (i^*, j^*). Since we have assumed our claim to be false, (i^*, j^*) is not strongly i-influential for any realization

$$(\rho_{i^1 j^1}(t), \rho_{i^2 j^2}(t), \ldots, \rho_{i^l j^l}(t)),$$

or the allocation of agent i in future rounds is the same whether $\rho_{i^*j^*}$ is 0 or 1 for every given $(\rho_{i^1 j^1}(t), \rho_{i^2 j^2}(t), \ldots, \rho_{i^l j^l}(t))$. This means that the allocation of agent i is the same in future rounds for all realizations $(\rho_{i^1 j^1}(t), \rho_{i^2 j^2}(t), \ldots, \rho_{i^l j^l}(t), \rho_{i^* j^*}(t))$. But this contradicts the fact that

$$\{(i^1, j^1), (i^2, j^2), \ldots, (i^l, j^l), (i^*, j^*)\},$$

is the set of i-influential slot–agent pairs in round t. This proves our claim. □

In our characterization of truthfulness under various settings, we show that a truthful allocation rule A must be type-I separated. Though the proofs look similar, there are subtle differences in each of the following subsections. In our proofs, we start with the assumption that a truthful allocation rule A is not type-I separated. That is,

$$\exists b_i < b_i^+, b_{-i}, \rho, t, t', (i^*, j^*) \ni (i^*, j^*) \in N((b_i, b_{-i}), \rho, t, i) \text{ with influenced round } t' \text{ and } A_{i^*j^*}((b_i^+, b_{-i}), \rho, t) = 0. \quad (1)$$

Subsequently, we show that this leads to a contradiction in each of the subsections, implying the necessity of type-I separatedness.

Unknown and unconstrained CTRs

In this setting, we do not assume any previous knowledge of the CTRs, although we do assume that such CTRs exist. Here, we show that any mechanism that is truthful under such a setting must follow some very rigid restrictions on its allocation rule.

Definition 11. (*strong pointwise monotonicity*). *An allocation rule is said to be strongly point-wise monotone if it satisfies: For any fixed* (b_{-i}, ρ), *if an agent i with bid* b_i *is allocated a slot j in round t, then* $\forall_{b_i}^+ > b_i$, *she is allocated the same slot j in round t. That is if the agent i receives a slot in round t, then she receives the same slot for any higher bid. For any lower bid, either she may receive the same slot or may lose the impression.*

A strongly pointwise monotone allocation rule is also a weakly pointwise monotone. However, strongly pointwise monotone is much stronger notion.

Theorem 1. *Let* (A, P) *be a deterministic, non-degenerate mechanism for the MAB, multi-slot sponsored search auction, with unconstrained and unknown* μ_{ij}. *Then, mechanism* (A, P) *is DSIC iff A is strongly pointwise monotone and type-I separated. Further, the payment scheme is given by,*

$$P_i((b_i, b_{-i}), \rho) = b_i C_i((b_i, b_{-i}), \rho) - \int_0^{b_i} C_i((x, b_{-i}), \rho)\mathrm{d}x.$$

Proof. The proof is organized as follows. In step 1, we show the necessity of the payment structure. In step 2, we show the necessity of strong pointwise monotonicity. As there are similarities in the proof of necessity of type-I separatedness in this theorem as well as in Proposition 1 (see later), we defer the proof to the Appendix. Finally in step 3, we prove that the above payment scheme in conjunction with strong pointwise monotonicity and type-I separatedness imply that the mechanism is DSIC.

Step 1: The utility structure for each agent $i \in N$ is

$$U_i(v_i, (b_i, b_{-i}), \rho) = v_i C_i((b_i, b_{-i}), \rho) - P_i((b_i, b_{-i}), \rho).$$

The mechanism is DSIC *iff* it is the best response for each agent to bid truthfully. That is, by bidding truthfully, each agent's utility is maximized. Under non-degenerate allocation rule assumption, by Myerson' theorem for single parameter agents, we need $(\mathrm{d}C_i/\mathrm{d}b_i) \geq 0$, which is the clickwise monotonicity condition.

And for (A, P) to be DSIC, we need

$$P_i((b_i, b_{-i}), \rho) = b_i C_i((b_i, b_{-i}), \rho) - \int_0^{b_i} C_i((x, b_{-i}), \rho)\mathrm{d}x + P_i((0, b_{-i}), \rho) \text{ and } \frac{\mathrm{d}C_i}{\mathrm{d}b_i} \geq 0 \quad \forall((b_i, b_{-i}), \rho). \quad (2)$$

For a mechanism to be normalized we need $P_i((0, b_{-i}), \rho) = 0$. And hence the necessity of the payment structure specified in the theorem.

Step 2: We first prove the necessity of strong pointwise monotonicity by contradiction. We have seen from eq. (2) that $(\mathrm{d}C_i/\mathrm{d}b_i) \geq 0$, $\forall((b_i, b_{-i}), \rho)$ is necessary for DSIC of A. We show that if A is not strongly pointwise monotone, then there exists some allocation and realization ρ for which $(\mathrm{d}C_i/\mathrm{d}b_i) < 0$. If A is not strongly pointwise monotone, there exists $(b_i, b_i^+, b_{-i}, \rho, t) \ni$

$$A_{ij_1}((b_i, b_{-i}), \rho, t) = 1 \text{ and } A_{ij_2}((b_i^+, b_{-i}), \rho, t) = 1, \quad \text{where } j_1 \neq j_2. \quad (3)$$

Over all such counter-examples, choose the one with the minimum t. By this choice, we ensure that in this example $\forall t' < t$, we have $A_{ij}(b_i, t') = A_{ij}(b_i^+, t')$. The only difference occurs in round t. Now, consider the game instance where $\rho_{ij_1}(t) = 1$, $\rho_{ij_2}(t) = 0$, $\rho_{ij}(\tau) = 0$, $\forall \tau > t$. The occurrence of such ρ has non-zero probability. Now, under (b_{-i}, ρ), agent i has the same allocation and the same number of clicks until round $(t - 1)$ independent of whether she bids b_i or b_i^+. However, in round t with bid b_i, she receives a click and with bid b_i^+ she does not, implying for this case that $(\mathrm{d}C_i/\mathrm{d}b_i) < 0$. This violates the click monotonicity requirement. So, strong pointwise monotonicity is indeed a necessary condition for truthful implementation of MAB mechanisms under this setting.

Step 3: Finally, we show that strong pointwise monotonicity and type-I separatedness are sufficient conditions for clickwise monotonicity and computability of the payments and hence for truthfulness. Suppose A is a strongly pointwise monotone and type-I separated allocation rule. So, it clearly satisfies the clickwise monotonicity. Now, by the type-I separatedness condition, we already have all the information required to calculate the allocation of agent i in every round for every bid $x < b_i$. This is because for bids $x < b_i$, by the strong pointwise monotonicity condition, in each round t, either agent i keeps the same slot she had in the observed game instance $((b_i, b_{-i}), \rho)$ or loses the impression altogether, that is, does not get a click. Hence, we have all the information required to compute $P_i((b_i, b_{-i}), \rho)$ according to eq. (2). This completes the sufficiency part of the theorem. □

Implications of strong pointwise monotonicity: For a given round t, if an agent i is allocated a slot j, then by the definition of strong pointwise monotonicity she receives the same slot for any higher bid that she places. If she lowers her bid, she may either retain the slot j, or lose the impression entirely. This leads to the strong restriction that an agent's bid can only decide whether or not she obtains an impression, and not which slot she actually gets. As we shall show below, this restriction has serious implications on the regret incurred by any truthful mechanism.

Regret estimate: In the single slot case it is a known result[11] that the worst case regret is $\Theta(T^{2/3})$. So, for the multi-slot case, the regret is $\Omega(T^{2/3})$. We show here that the worst case regret generated in the multi-slot general setting by a truthful mechanism is in fact $\Theta(T)$. We show this for the two-slot, three-agent case with an intuitive argument, which can be generalized.

Consider a setting with two slots and three competing agents, that is $m = 2$, $k = 3$. Let the agents be A_1, A_2 and A_3. By Theorem 1, any truthful mechanism has to be strongly pointwise monotone. That is, in any round, the bids of the agents only determine which agents will be displayed and not the slots they obtain.

Suppose, A_3's bid $b_3 < \min(b_1, b_2)$ in addition to having low CTRs. In this case, any mechanism that grants A_3 an impression $\Theta(T)$ times, will have regret $\Theta(T)$.

So, we can assume that A_3' s ad gets an impression for a very small number of times when compared with T. Thus, ads by A_1 and A_2 will appear $\Theta(T)$ times. In each round, A_1 will get either slot 1 or slot 2 independent of her bid, while the other slot is assigned to A_2.

In any strongly pointwise monotone mechanism, either A_1 is assigned a slot 1 $\Theta(T)$ times or slot 2.

Without loss of generality, we assume that A_1 is assigned slot 1 $\Theta(T)$ times. So, the allocation (slot 1, slot 2) $\leftrightarrow$ (A_1, A_2) is made $\Theta(T)$ times. Consider a game instance where this is not the welfare maximizing assignment, that is, the relation $(\mu_{11}b_1 + \mu_{22}b_2) < (\mu_{12}b_1 + \mu_{21}b_2)$ holds true. Since the slot allocation does not depend on the individual bids, such an instance can occur. In such a setting (A_2, A_1) would have been optimal assignment. As a result, each round having the allocation (A_1, A_2) incurs constant non-zero regret. Since such an allocation occurs $\Theta(T)$ times, the mechanism has a worst case regret of $\Omega(T)$. However, there are T rounds, the regret is $O(T)$. Hence any truthful mechanisms under the unrestricted CTR setting exhibit a high $\Theta(T)$ regret. □

Since the strong monotonicity condition places such a severe restriction on A and also leads to a very high regret, in the following sections we explore some relaxations on the assumption that μ_{ij}'s are unrelated. With such settings which are in fact practically quite meaningful, we are able to prove more encouraging results.

Higher slot click precedence

This setting is similar to the general one discussed above in that we do not assume any knowledge about the CTRs. However, we impose a restriction on the realization ρ that it follows *higher slot click precedence* defined below.

Definition 12. *A realization ρ is said to follow higher slot click precedence if* $\forall i \in K$, $\forall t = 1, 2, \ldots, T$,

$$\rho_{ij_1}(t) = 1 \Rightarrow \rho_{ij_2}(t) = 1 \ \forall j_2 < j_1.$$

Higher slot click precedence implies that if an agent i obtains a click in slot j_1 in round t, then in that round, she receives a click in any higher slot j_2. This assumption is in general valid in the real world since any given user (fixed by round t) who clicks on a particular ad when it is displayed in a lower slot would definitely click on the same ad if it was shown in a higher slot.

We show, under this setting, that *weak pointwise monotonicity* and *type-I separatedness* are necessary conditions for truthfulness. They are, however, not sufficient conditions. Clearly, strong pointwise monotonicity and type-I separatedness will still be sufficient conditions. A weaker sufficient condition for truthfulness under this setting is still elusive.

Implications of the assumption: Observe that a slot–agent pair (i, j) is influential in some round t only if changing the realization in the entry $\rho_{i,j}(t)$ for some realization ρ results in a change in allocation in some future round. Crucial to the influentiality is the fact that $\rho_{ij}(t)$ can change.

Now, consider the following situation: it has been observed that in the game instance $((b_i, b_{-i}), \rho)$, we have $\rho_{ij_1}(t) = 0$, where agent i obtains slot j_1 in round t. We are interested in the game instance $((x, b_{-i}), \rho)$ where agent i gets slot $j_2 > j_1$, where $x < b_i$ and in knowing whether (i, j_2) is an influential pair in round t for some influenced agent. Now, since $\rho_{ij_1} = 0$ and $j_1 < j_2$, by our defining assumption, we conclude that $\rho_{ij_2}(t) = 0 \ \forall x < b_i$. Hence, our mechanism knows that in all the relevant cases, the realization in the given slot–agent pair never changes. Hence, (i, j_2) cannot be an influential pair for any $j_2 > j_1$ in round t. We will use this observation in the proof of necessity characterization.

Proposition 1. *Consider the setting in which realization ρ follows higher slot click precedence. Let (A, P) be a deterministic non-degenerate DSIC mechanism for this setting. Then the allocation rule A must be weak pointwise monotone and type-I separated. Further, the payment scheme is given by,*

$$P_i(b_i, b_{-I}; \rho) = b_i C_i(b_i, b_{-I}; \rho) - \int_0^{b_i} C_i(x, b_{-i}; \rho)\mathrm{d}x .$$

Proof. The proof for the payment scheme is identical to that in Theorem 1. Here, we prove the necessity of weak pointwise monotonicity and for type-I separatedness, we refer it to Appendix.

We prove the necessity of weak pointwise monotonicity, in a very similar fashion to that of the necessity of strong pointwise monotonicity in Theorem 1. The crucial difference is, while constructing ρ, we have to ensure that it satisfies the higher order click precedence. Suppose A is truthful but not weakly pointwise monotone, that is, $\exists$ $(b_i, b_i^+, b_{-i}, \rho, t)$ and $A_{ij_1}((b_i, b_{-i}), \rho, t) = 1$ and $A_{ij_2}((b_i^+, b_{-i}), \rho, t) = 1$ for some $j_1 < j_2$. Over all such examples, choose the one with the least t. By this choice, we ensure that in this example, $\forall t' < t$, we have $A_{ij}(b_i, t') = A_{ij}(b_i^+, t')$. The only difference occurs in round t. Now, consider the game instance where $\rho_{ij_1}(t) = 1$ and $\rho_{ij_2}(t) = 0$. Such realization has a non-zero probability of occurrence. Now, under (b_{-i}, ρ), agent i gets the same allocation and the same number of clicks until round $(t-1)$ independent of whether she bids b_i or b_i^+. However, in round t with bid b_i she gets a click and with bid b_i^+ she does not, implying for this case that $(\mathrm{d}C_i/\mathrm{d}b_i) < 0$. This leads to a contradiction. So, weak pointwise monotonicity is a necessary condition. If A is not strongly pointwise monotone, it does not violate clickwise monotonicity. That is, for truthful A, it may possible that, $A_{ij_1}((b_i, b_{-i}), \rho, t) = 1$ and $A_{ij_2}((b_i^+, b_{-i}), \rho, t) = 1$ where $j_2 < j_1$. Thus, for A to be truthful, strong pointwise monotonicity may not be necessary. □

When CTR pre-estimates are available

In this setting, we assume the existence of some previous database or pre-estimate of CTR values, but *no restriction on ρ*. That is, $\mu_{ij} = X_{ij}/Y_{ij}$, where X_{ij} is the number of clicks obtained by agent i in slot j out of the Y_{ij} times she obtained the slot j over all past auctions. Here, in general, $\mu_{i1} \geq \mu_{i2} \geq \cdots \geq \mu_{im}$. For our characterization, we assume that each $\mu_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{im})$ is known to the agent i and the auctioneer.

In this setting, the auctioneer uses explorative rounds to improve his estimate of the CTRs and updates the database. Then, he makes use of his new knowledge of the CTRs in the exploitative rounds. The payment scheme, however, only makes use of the old CTR matrix. Under this scheme, we derive the conditions required for a mechanism to be truthful in expectation over μ, defined as follows.

Definition 13. (*truthful in expectation*). *A mechanism is said to be truthful in expectation over μ, the CTR pre-estimate, if each of the agents believes that the number of clicks she obtains is indeed $\sum_t\sum_j(\mu_{ij}A_{ij}(b, \rho, t))$, which is the number of clicks she will obtain if the CTR pre-estimate is perfectly accurate.*

Fairness: For this characterization, we need the notion of fair allocation rules, as defined below.

Definition 14. (*fair allocation*). *Consider two game instances $((b_i, b_{-i}), \rho)$ and $((b_i', b_{-i}), \rho)$ having the same slot–agent–round triplets, (i', j', t') as strongly i-influential. Let (i^*, j^*, t) be such a triplet with the smallest t' in which i is influenced. Consider the realization ρ' differing from ρ only in this influential element $\rho_{i^*j^*}(t)$. Then, the allocation rule A is said to be fair if for every such pair of games it happens that*

$$\sum_j \mu_{ij} A_{ij}((b_i, b_{-i}), \rho, t') \geq \sum_j \mu_{ij} A_{ij}((b_i, b_{-i}), \rho', t')$$

$$\Leftrightarrow \sum_j \mu_{ij} A_{ij}((b_i', b_{-i}), \rho, t') \geq \sum_j \mu_{ij} A_{ij}((b_i', b_{-i}), \rho', t').$$

The intuition behind *fair allocations* is that changing the realization *only* in a fixed strongly i-influential slot generally changes agent i's allocation in a predictable fashion independent of her own bid, either improving her slot or worsening it in the earliest influenced round, irrespective of the allocation or realization in the rest of the game. For example, if agent i's chief competitor agent, i', is strongly i-influential, then i' not getting a click in the influential round will generally mean that agent i will go on to get a better slot than if agent i' got a click, independent of b_i.

Truthfulness characterization: Here, the expected utility (see note 1) for the agent i,

$$U_i(v_i, b, \rho) = \left(\sum_{t=1}^{T} \sum_{j=1}^{m} \mu_{ij} A_{ij}(b, \rho, t) v_i \right) - P_i(b, \mu). \qquad (4)$$

Proposition 2. *Let (A, P) be a normalized mechanism under this setting. Then, the mechanism is truthful in* *expectation over μ iff A is weakly pointwise monotone and the payment rule is given by*

$$P_i(b, \mu) = \sum_{t=1}^{T}\sum_{j=1}^{m} \mu_{ij}\{b_i A_{ij}(b, \mu, t) - \int_0^{b_i} A_{ij}(x, b_{-i}, \mu, t)\mathrm{d}x\},$$

and payments are computable.

Proof. In step 1, we prove the necessity and sufficiency of the payment structure. For the mechanism to be implemented, we need to compute the payments of all the agents uniquely. That is, P_is need to be computable for all agents i. In step 2, we show weak pointwise monotonicity is equivalent to the second-order condition which is clickwise monotonicity in the context of this article.

Step 1: The expected utility of an agent i is given by eq. (4). Under non-degeneracy, by Myerson's theorem we get for (A, P) to be truthful, the payment structure should be,

$$P_i(b, \mu) = \sum_{t=1}^{T}\sum_{j=1}^{m} \mu_{ij}\{b_i A_{ij}(b, \mu, t) - \int_0^{b_i} A_{ij}(x, b_{-i}, \mu, t)\mathrm{d}x\},$$

and

$$\forall i, \quad \sum_t\sum_j \mu_{ij}\frac{\mathrm{d}A_{ij}}{\mathrm{d}b_i} \geq 0. \tag{5}$$

Step 2: We show, eq. (5) $\Leftrightarrow$ *weak pointwise monotonicity.*

(i) It is obvious that *weak pointwise monotonicity* $\Rightarrow$ $\sum_t\sum_j \mu_{ij}$ $(\mathrm{d}A_{ij}/\mathrm{d}b_i) \geq 0$. An increase in b_i under a weakly pointwise monotone A would result in a better slot allocation for agent i. This in turn, would result in an increase in $\sum_j\mu_{ij} A_{ij}$ in each round.

(ii) Now we prove the converse. Suppose A is not weakly pointwise monotone. That is, $\exists i, b_i, b_i^+, b_{-i}, \rho, \mu, \ni$ $A_{ij}(b_i, b_{-i}, \rho, \mu, t) = 1$ and $A_{ij'}(b_i^+, b_{-i}, \rho, \mu, t) = 1$, where $j' > j$. Consider the smallest such t. Allocation in this round does not depend upon the realization of this round or of future rounds. We consider the instance of the game where $\rho_{ij}(t) = 1$ and $\rho_{ij'}(t) = 0$ and t is the last round. Such an instance has a non-zero probability and for this instance, $\sum_t \mathrm{d}/\mathrm{d}b_i \ \sum_j \mu_{ij}A_{ij} < 0$. This proves the equivalence claim. □

Note, it is crucial that each μ_{ij} is a previously known constant and cannot be defined as $\mu_{ij} = X_{ij}/Y_{ij}$ based on the clicks in the current T rounds post facto. If we do so, X_{ij}/Y_{ij} can change with the allocation of agent i in a particular game and hence, μ_{ij} would become a function of b_i and the mechanism would be no longer truthful.

For truthful implementation, the payments need to be computable and computing the payments may involve the unobserved part of ρ. In the next theorem, we show that type-I separatedness is necessary and type-II separatedness is sufficient for computation of these payments. So, along with the computation of payments and the above proposition, we get the following.

Theorem 2. *Let (A, P) be a mechanism for this stochastic multi-round auction setting where A is a non-degenerate, deterministic and fair allocation rule. Then, (A, P) is truthful in expectation over μ if A is weakly pointwise monotone and type-II separated and the payment scheme is given by,*

$$P_i(b, \rho) = \sum_{t=1}^{T}\sum_{j=1}^{m} \mu_{ij}\{b_i A_{ij}(b, \rho, t) - \int_0^{b_i} A_{ij}(x, b_{-i}, \rho, t)\mathrm{d}x\}.$$

Also, if a mechanism (A, P) is truthful, then it is weakly pointwise monotone, type-I separated, the payment is given as above and is computable.

Proof (*see note 2*). This setting/characterization works best with old advertisers who have already taken part in a large number of auctions. As we already have proved Proposition 2, we just need to show that type-I separatedness is in fact a necessary and type-II separatedness is sufficient condition for the computability of payments, that is, for each agent i, computability of

$$\sum_{j=1}^{m} \mu_{ij}\int_0^{b_i} A_{ij}(x, b_{-i}, \rho, t)\mathrm{d}x.$$

Step 1: We first provide the proof for the sufficiency of type-II separatedness. Suppose A is type-II separated. The mechanism observes and knows all allocations and the observed realization for the game instance carried out with the original bid vector (b_i, b_{-i}). Specifically, it knows $N((b_i, b_{-i}), i, \rho, t)$ for all rounds t and the respective realizations in these slots. Now, in the game instance (x, b_{-i}) where $x \leq b_i$, by type-II separatedness, we have $N((x, b_{-i}), i, \rho, t) \subseteq N((b_i, b_{-i}), i, \rho, t)$. This means that the allocation in i-influential slots for game instance $((x, b_{-i})t)$ is a subset of that in observed game instance $((b_i, b_{-i}), \rho)$. So, the mechanism already knows all the click information in the i-influential slots for the game instance $((x, b_{-i}), \rho)$. Since the payment scheme is only interested in the allocation of agent i, the realization in the unobserved slots is unimportant and can be assumed arbitrarily. Thus, the mechanism has complete information to compute $P_i((b_i, b_{-i}), \rho, t)$.

Step 2: Next, we prove the necessity of type-I separatedness by contradiction. That is, we assume eq. (1) is true. Consider a *complete* realization $\rho(t)$ in round t for which (i^*, j^*) is strongly i-influential (such a realization exists by our previous claim 1) and construct the two complete realizations ρ and ρ' from $(\rho(1), \rho(2), \ldots, \rho(t-1), \rho(t))$ which only differ in $\rho_{i^*j^*}(t)$. Over all choices of counter-examples $(b_i, t, \rho(t), i^*, j^*)$, we choose the one which has the smallest influenced round t'. Now, we compare the payment that the mechanism has to make for this game instance at the end of t' rounds under the two different realizations ρ and ρ'.

Let $\varphi \in \{\rho, \rho'\}$. By the strong i-influence of (i^*, j^*, t), the agent i gets different allocations in round t' under the different realizations ρ and ρ'. This implies, $\sum_j \mu_{ij} A_{ij}((b_i, b_{-i}), \rho, t') \neq \sum_j \mu_{ij} A_{ij}((b_i, b_{-i}), \rho', t')$.

Without loss of generality,

$$\sum_j \mu_{ij} A_{ij}((b_i, b_{-i}), \rho, t') > \sum_j \mu_{ij} A_{ij}((b_i, b_{-i}), \rho', t'), \quad (6)$$

(or agent i gets a higher slot under realization ρ than ρ').

By the *non-degeneracy* of A, there exists a finite interval of bids about b_i such that for every bid x in this interval,

$$A_{ij}((x, b_{-i}), \varphi, t') = A_{ij}((b_i, b_{-i}), \varphi, t') \quad \forall j. \quad (7)$$

Suppose $x' \in (0, b_i^+)$ is another bid such that the same slot–agent–round set (i^*, j^*, t) is strongly i-influential with the same influenced round t' for the game $((x', b_{-i}), \rho)$. Then by the fairness of A,

$$\sum_j \mu_{ij} A_{ij}((x', b_{-i}), \rho, t') \geq \sum_j \mu_{ij} A_{ij}((x', b_{-i}), \rho', t'). \quad (8)$$

From eqs (6)–(8) and using the fact that t' is the smallest influenced round that is strongly i-influenced by the bit $\rho_{i^*j^*}(t)$ which is the only differing bit between ρ and ρ', we can see that $\forall x' \in (0, b_i^+)$

$$\sum_j \mu_{ij} A_{ij}((x', b_{-i}), \rho, t') \geq \sum_j \mu_{ij} A_{ij}((x', b_{-i}), \rho', t'). \quad (9)$$

and $\exists$ a finite interval X around bid b_i such that $\forall x \in X$, we have

$$\sum_j \mu_{ij} A_{ij}((x, b_{-i}), \rho, t') > \sum_j \mu_{ij} A_{ij}((x, b_{-i}), \rho', t'). \quad (10)$$

From eqs (9) and (10), and the fact that agent i's allocation is the same under both realizations ρ and ρ' until round t' (from the smallest influenced round choice), we conclude that,

$$\sum_{t=1}^{t'} \int_0^{b_i^+} \sum_j \mu_{ij} A_{ij}((x, b_{-i}), \rho, t)\, \mathrm{d}x$$

$$> \sum_{t=1}^{t'} \int_0^{b_i^+} \sum_j \mu_{ij} A_{ij}((x, b_{-i}), \rho', t)\, \mathrm{d}x.$$

Additionally, we can assume that there are no clicks after round t'. As a result, we have $P_i((b_i^+, b_{-i}), \rho) \neq P_i((b_i^+, b_{-i}), \rho')$. However, the mechanism cannot distinguish between the two realizations ρ and ρ' as the only differing bit $\rho_{i^*j^*}(t)$ is unobserved. Hence, the mechanism fails to assign a unique payment to agent i. This is a consequence of our initial assumption, eq. (1). Thus if A is not type-I separated, the payments are not computable. This completes the proof. □

When CTR is separable

In the previous setting we assumed that some pre-estimate on the CTR matrix $[\mu_{ij}]$ existed. In real-world applications, however, it is very often the case that the slot-dependent probabilities are known while the agent-dependent probabilities are unknown. To leverage this fact, we make a widely accepted assumption: we assume that the click probability due to the slot is independent of the click probability due to the agent. That is, we assume that $\mu_{ij} = \alpha_i \beta_j$, where α_i is the click probability associated with agent i and β_j is the click probability associated with slot j. We also assume that the vector $\beta = (\beta_1, \beta_2, \ldots, \beta_m)$ is common knowledge. In general $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_m$. Here, any mechanism will use the explorative rounds to try to learn the values of α_i as accurately as possible.

Let Y_{ij} denote the number of times that agent i obtains the impression for slot j, and X_{ij} denote the corresponding number of times she obtains a click. Then, we define

$$\alpha_i' = \mathrm{avg}_j \left\{ \frac{1}{\beta_i} \frac{X_{ij}}{Y_{ij}} \right\} \text{ and } \mu_{ij}' = \alpha_i' \beta_j.$$

In this section, we assume $\mathrm{d}\alpha_i'/\mathrm{d}b_i = 0$ or that α_i' does not change with bid b_i. We are justified in making this assumption since α_i' is a good estimate of α_i, which is independent of which slot agent i obtains how many times. By changing her bid b_i, agent i can only alter her allocations which should not predictably or significantly affect α'. It is trivial to see that

$$\frac{\mathrm{d}\alpha_i'}{\mathrm{d}b_i} = 0 \Rightarrow \frac{\mathrm{d}\mu_{ij}'}{\mathrm{d}b_i} = \frac{b_j \mathrm{d}\alpha_i'}{\mathrm{d}b_i} = 0.$$

We model truthfulness based on the utility gained by each agent in expectation over this μ_{ij}'. That is, utility to an agent i is given by eq. (4), with μ being replaced by μ'. With the above setup, it can be easily seen that truthfulness mechanisms under this setting have the same characterization as the truthful mechanisms with a pre-estimate of CTR.

Theorem 3. *Let* (A, P) *be a mechanism for the stochastic multi-round auction setting where* A *is a non-*

degenerate, deterministic and fair allocation rule. Then (A, P) *is truthful in expectation over* μ' *if* A *is weakly pointwise monotone and type-II separated and the payment scheme is given by*,

$$P_i(b, \rho) = \sum_{t=1}^{T}\sum_{j=1}^{m}\mu'_{ij}\left\{b_i A_{ij}(b, \rho, t) - \int_0^{b_i} A_{ij}(x, b_{-i}, \rho, t)\mathrm{d}x\right\}.$$

Also, if a mechanism (A, P) *is truthful, then it is weakly pointwise monotone, type-I separated, the payment is given as above and is computable.*

Proof. This theorem can be proven using similar arguments as used in the proof of Theorem 2, with μ being replaced by μ'. □

Experimental analysis

Since the single-slot setting is a special case of the multi-slot setting, we obtain $\Omega(T^{2/3})$ as a lower bound for the regret incurred by a truthful multi-slot sponsored search mechanism.

We have characterized truthful MAB mechanisms in various settings in the previous section. However, we have not studied MAB mechanisms in multi-slot auctions for regret estimation in such mechanisms (except the $\Theta(T)$ worst-case bound we showed for the unconstrained case earlier). In this section, we present a brief experimental study on the regret of an truthful MAB mechanism for multi-slot sponsored search auction under separable CTR case.

For our study, we have picked a simple mechanism belonging to the separable CTR case. In the simulation,

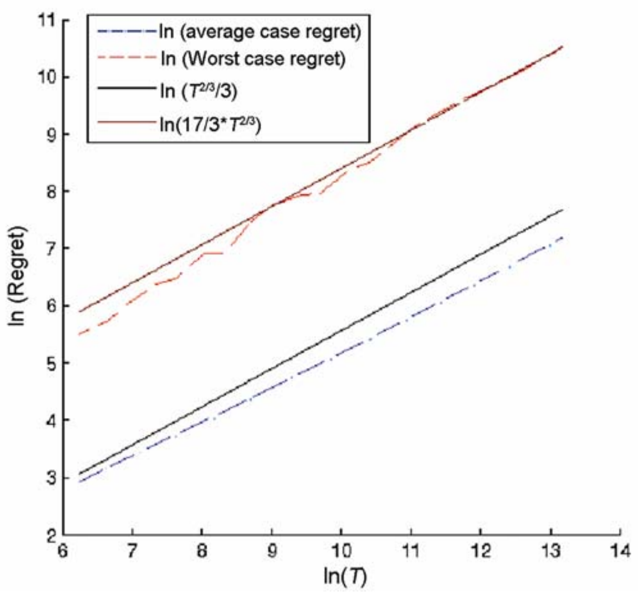

Figure 3. Average case regret and worst case regret in a logarithmic scale.

we displayed the agents in the available slots in a round-robin fashion for the first $T^{2/3}$ rounds. Then, we used the observed information on the clicks to estimate the μ_{ij} values. The payments were computed according to Theorem 3.

We performed simulations for various T values with $k = 4$ and $m = 2$. For a fixed T, we generated $100T$ different instances and estimated the average case as well worst case regrets. In each instance, we generated CTRs and bids randomly. Figure 3 depicts ln(worst case regret) and ln(average case regret). It is observed that ln(worst case regret) is closely approximated by $\ln((17/3)T^{2/3})$, while ln(average case regret) is closely approximated by $\ln((1/3)T^{2/3})$, clearly showing that the worst case regret is $\Omega(T^{2/3})$ and the average case regret is upper bounded by $O(T^{2/3})$.

Conclusion

In this article, we have provided characterizations for truthful multi-armed bandit mechanisms for various settings in the context of multi-slot pay-per-click auctions, thus generalizing the work of Babaioff *et al.*[11] in a non-trivial way. The first result we proved is a negative result which states that under the setting of unrestricted CTRs, any strategyproof allocation rule is necessarily strongly pointwise monotone. We also showed that every strategyproof mechanism in unrestricted CTR setting will have $\Theta(T)$ regret. By weakening the notion of unrestricted CTRs, we were able to derive a larger class of strategyproof allocation rules. Our results are summarized in Table 1.

In the auctions that we have considered, the auctioneer cannot vary the number of slots he wishes to display. One possible extension of this work could be in this direction, that is, the auctioneer can dynamically decide the number of slots for advertisements. We assume that the bidders bid their maximum willingness to pay at the start of the first round and they would not change their bids till T rounds. Another possible extension would be to allow the agents to bid before every round. We could also be exploring the cases where the bidders have budget constraints.

Appendix. Proofs

We show the necessity of type-I separatedness for Theorem 1 and Proposition 1. We prove the necessity of the type-I separatedness condition by contradiction. That is we assume eq. (1). Over all such possible counter-examples of (b_i, b_i^+, $\underline{b}_{-i}$, ρ, t, t'), choose the one with the least t'. We can assume, $\rho_{ij}(\tau) = 0$, $\forall\, \tau > t'$, as the clicks in future rounds do not affect decisions in the current round. Let ρ' be the realization that differs from ρ only in bit $\rho_{i^*j^*}(t)$. We show that under these two realizations the payment charged to the agent i differs at bid b_i^+. However, since (i^*, j^*) is not part of the allocation in round t under the bid b_i^+, the difference between ρ and ρ' is not observed by the mechanism.

Agent i's allocation and click information differ only in round t' under the two different realizations ρ and ρ', by minimality of t'. Now, for the proof of Theorem 1, by strong monotone property agent i is either displayed in a particular slot, say j or not displayed at all. (i^*, j^*) being i-influential, in round t', agent i's allocation differs in ρ and ρ'. Let,

$$0 = A_{ij}((b_i, b_{-i}), \rho, t') < A_{ij}((b_i, b_{-i}), \rho', t') = 1.$$

Otherwise we can swap ρ and ρ'. Now,

$$A_{ij}((x, b_{-i}), \rho, t') \le A_{ij}((x, b_{-i}), \rho', t') \quad \forall x < b_i^+. \qquad (11)$$

If above claim is not true we can get violation of strong pointwise monotone necessity either at ρ or ρ'. Because of non-degeneracy, there exists interval X such that

$$A_{ij}((x, b_{-i}), \rho, t') < A_{ij}((x, b_{-i}), \rho', t') \quad \forall x \in X.$$

Hence $P_i((b_i^+, b_{-i}), \rho) < P_i((b_i^+, b_{-i}), \rho)$ proving the claim.

Now, for the proof of Proposition 1, (i^*, j^*) being i-influential, in round t', agent i's allocation differs in ρ and ρ'. Let

$$A_{ij_1}((b_i, b_{-i}), \rho, t') = 1 \text{ and } A_{ij_2}((b_i, b_{-i}), \rho', t') = 1.$$

Here since the two differ only in $j_1 \ne j_2$, without loss of generality, let $j_1 < j_2$ (or j_1 be the better slot, since it is possible that one of the realizations may lead to no slot allocation). Now, similar to observation in eq. (11), we can argue that, for all $x < b_i^+$, the agent i is allocated a better slot than j_2 under realization ρ. Due to higher slot click precedence and non-degeneracy of allocation rule, $P_i((b_i^+, b_{-i}), \rho) < P_i((b_i^+, b_{-i}), \rho)$.

Thus we have seen that, if the allocation rule is not type-I separated in the context of either Theorem 1 or Proposition 1, the mechanism cannot uniquely determine the truthful payments for the agents. Hence, type-I separatedness is necessary for truthful implementation of a clickwise monotone allocation rule. □

Notes

1. The characterization in this section would hold even if μ_{ij} are arbitrary weights. However, while using arbitrary weights, mechanism may charge some agents more than their actual willingness to pay. Also regret in the revenue, that is loss in the revenue to the search engine will be trivially $\Theta(T)$.
2. The idea for our proof is similar to that in the characterization of the single-slot case[11], however, the details are different.

1. Seda, C., *Search Engine Advertising: Buying Your Way to the Top to Increase Sale (Voices that Matter)*, New Riders Press, New York, 2004.
2. Hoffman, D. L. and Novak, T. P., Advertising pricing models for the world wide web. In *Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property* (eds Kahin, B. and Varian, H. R.), The MIT Press, 2000.
3. Adams, R., *www.Advertising: Advertising and Marketing on the World Wide Web (Design Directories)*, Watson-Guptill Publications, New York, 2003.
4. Zeff, R. L., *Advertising on the Internet*, John Wiley, 1999.
5. Narahari, Y., Garg, D., Rama Suri, N. and Prakash, H., *Game Theoretic Problems in Network Economics and Mechanism Design Solutions*, Advanced Information and Knowledge Processing Series, Springer, London, 2009.
6. Robbins, H., Some aspects of the sequential design of experiments. *Bull. Am. Math. Soc.*, 1952, **58**, 527–535.
7. Gittins, J. C., Glazebrook, K. and Weber, R., *Multi Armed Bandit Allocation Indices*, John Wiley, 2011.
8. Auer, P., Cesa-Bianchi, N. and Fischer, P., Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.*, 2002, **47**, 235–256.
9. Lai, T. L. and Robbins, H., Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.*, 1985, **6**, 4–22.
10. Bergemann, D. and Valimaki, J., Bandit problems. Cowles Foundation Discussion Paper, Number 1551, 2006.
11. Babaioff, M., Sharma, Y. and Slivkins, A., Characterizing truthful multi-armed bandit mechanisms. In EC09: Proceedings of the 10th ACM Conference on Electronic Commerce, Stanford, California, 2009, pp. 79–88.
12. Devanur, N. R. and Kakade, S. M., The price of truthfulness for pay-per-click auctions. In EC09: Proceedings of the 10th ACM Conference on Electronic Commerce, 2009, pp. 99–106.
13. Babaioff, M., Kleinberg, R. D. and Slivkins, A., Truthful mechanisms with implicit payment computation. In Proceedings 11th ACM Conference on Electronic Commerce, 2010, pp. 43–52.
14. Zoeter, O., On a form of advertiser cheating in sponsored search and a dynamic-VCG solution. In Proceedings of TROA, 2008.
15. Gonen, R. and Pavlov, E., An incentive-compatible multi-armed bandit mechanism. In PODC'07, New York, 2007, pp. 362–363.
16. Gatti, N., Lazaric, A. and Trovò, F., A truthful learning mechanism for contextual multi-slot sponsored search auctions with externalities. In Proceedings 13th ACM Conference on Electronic Commerce, 2012, pp. 605–614.
17. Archer, A. and Tardos, É., Truthful mechanisms for one-parameter agents. In Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science, Washington, DC, USA, pp. 482–491, 2001.
18. Myerson, R. B., Optimal auction design. *Math. Oper. Res.*, 1981, **6**, 58–73.