

# Optimal User Search

Nicole Immorlica, Northwestern University  
Mohammad Mahdian, Google  
Greg Stoddard, Northwestern University

Most research in sponsored search auctions assume that users act in a stochastic manner, ignoring potential strategic deliberations. We propose a simple model of user costs and benefits. We then show how to formulate the corresponding sponsored search environment as a Markov decision process. For certain sponsored-search mechanisms, we are able to derive the optimal (i.e., utility-maximizing) user search behavior. The optimal behavior is characterized by a threshold policy in which users view ads in a top-down manner and click on those whose perceived quality exceeds a fixed threshold.

## 1. INTRODUCTION

When a user submits a query to a search engine, the search engine displays a list of ‘algorithmic’ results alongside a ranked list of ‘sponsored’ results, or advertisements. These sponsored results are sold to advertisers via an auction. In a typical sponsored-search auction, advertisements are ranked based on two attributes for each advertiser, *value* and *quality*. Value is the advertiser’s maximum willingness to pay for a click on their ad and quality is usually interpreted as the probability that this advertiser will satisfy the search need of a user visiting that site. Search engine users then search through these advertisements in an effort to satisfy a search need. Given that users have a cost associated with each searching action (reading the small text of the ad, visiting a website), and a benefit for satisfying the search need, we ask the following question: What is the optimal search strategy for users?

Most prior work analyzing sponsored-search auctions assumes stochastic models for user search behavior. The early *separable click-through-rate* model and later *cascade* model assume the user clicks on each ad with an exogenous probability related to the ad itself, its rank in the list, and (in the case of the cascade model) the rank of other ads. These models enjoy tremendous success in explaining the welfare properties of sponsored-search auctions for advertisers in equilibria, and the resulting revenue for the search engine. However, neither model considers users as strategic agents, and thus forbids analyses that directly address user welfare.

We overcome this shortcoming by modeling users as strategic agents. To understand the model, let’s consider the setting in a bit more detail. A user submits a query, say “camera” for example, to a search engine in order to satisfy a search need, buying a new SLR perhaps. In response to the query, the search engine presents a ranked list of ads that it hopes will satisfy the user’s need. The user reads the ads, clicking on some subset of them, until he either satisfies his search need (finds a satisfactory camera), or abandons the search. The user has some value for satisfying his search need, and incurs costs in the search process. Reading the text of an advertisement in the ranking imposes a small viewing cost on the user, related to the time it takes to read and evaluate the ad. After reading, the user has an estimate of the quality of the ad, i.e., the probability the ad satisfies his search need. By incurring an additional (and larger) cost, the user can click on the advertisement, visit the advertiser’s website, and determine whether the ad does indeed satisfy his search need.

In this work, we present a framework for determining optimal user behavior based on the sponsored search environment, namely users’ benefits and costs, the distributions over advertisers’ qualities and values, and the sponsored-search mechanism, or ranking rule, used by the search engine. We assume that, given the quality and value distributions and ranking rule, users can determine the distribution of quality for each advertisement in the list (conditional on prior advertisements). Thus the optimization problem facing the user reduces to the following simple Markov decision process: there is a stream of coins

(advertisements). Each coin has a probability of heads (ad quality) that is drawn according to a known joint distribution. The user maintains a bag of coins, initially empty. At each step, the user can: 1) pay a small cost to purchase the next coin in the stream<sup>1</sup> and hence observe the realization of its probability of heads; 2) pay a larger cost to flip a coin in his bag and hence satisfy his search need (receiving the value of the search) if the coin flip is heads; 3) abandon the search.

We consider two special cases of the sponsored search environment. In the first case, the ranking rule sorts ads in order of quality. Thus the highest quality ad is the first ad that the user views. The optimal user behavior, as observed in prior work, is thus to read and click in a top-down manner until the expected quality of the next ad fails to exceed the clicking cost. In the second case, the ranking rule sorts ads in random order with respect to qualities. Our main result is that, for this setting, the following simple threshold policy is optimal: read ads one-by-one and click any ad that exceeds the computed threshold.

Although for general quality distributions we can only represent the threshold implicitly, for uniform distributions we can explicitly characterize the user value of the optimal policy in the two cases of interest. This allows us to compare the social welfare (sum of advertiser and user value) for these two cases. We show that, depending on the view and click costs, either rule can optimize social welfare. The intuition is that the first case offers the user high value, but generates few clicks for advertisers, whereas the second case imposes a larger search cost on the user but also generates more clicks, giving each clicked advertiser some positive welfare (equal to his value for a click).

The remainder of the paper is as follows. Section 3 formally introduces our sponsored-search model and the stochastic process we use to study user search behavior. Section 4 presents an application of this model to the primary setting considered in this paper. Section 5 gives background for Markov Decision Processes and section 6 uses this background to prove that a threshold policy is optimal for the setting in Section 4. Sections 7 and 8 contain discussions regarding welfare comparisons and future directions for this work.

## 2. PREVIOUS WORK

Early work in the sponsored-search literature was largely focused on characterizing equilibria bidding strategies for advertisers [Varian 2007; Edelman et al. 2005]. These early models did not allow any advertiser-specific effects on click-through-rate or externalities on CTR between ads. Later work proposed various stochastic models of user search behavior to develop a more complete theory on the role of advertiser-specific effects and externalities in click-through-rates [Aggarwal et al. 2008; Kempe and Mahdian 2008]. In the context of these stochastic models for user behavior, prior work also studies properties, like revenue and welfare, of various ranking rules [Lahaie and McAfee 2011; Lahaie and Pennock 2007; Gomes et al. 2009]. Our work can be viewed as a strategic analogue of this line of work; we first derive optimal strategic behavior for user search and then use these results to compare welfare properties of two simple ranking rules.

The work of Athey and Ellison [2009] is mostly closely related to our work. They consider a setting where users strategically search through displayed links and use this characterization to show a strong connection between user value and total social welfare. They assume that advertisers only receive value when they meet the search need of a consumer and thus quality and value are completely aligned in their model. In turn, their results pertain to equilibria

---

<sup>1</sup>Note by assuming users must always purchase coins in order, we are assuming in the sponsored search setting that users read ads in a top-down manner. This assumption is not without loss in all settings, although it will be in the setting we study in this paper. Exploring optimal behavior with an unrestricted reading order is interesting future work.

in which ads are sorted by quality. Our framework addresses a large class of ranking rules<sup>2</sup> and allows for more general relations between advertiser quality and value.

### 3. THE MODEL

#### 3.1. Sponsored-Search

An advertiser is defined by a quality and value pair  $(q, v)$  where an advertiser's quality is the probability that the advertiser's site satisfies a user's search need and an advertiser's value is the value he has for one visitor to his site. We assume that the  $(q, v)$  pair for each advertiser is drawn IID from distribution  $\mathcal{D}$ . A sponsored-search mechanism  $M$  produces an assignment of advertisers to slots based on the reports of (quality,value) pairs for all advertisers. For simplicity, we assume advertisers report their values truthfully. We note that the quality for an advertiser is different from its click-through-rate; quality is chance that a user will have his need satisfied by the ad while CTR is the expected number of clicks that an ad receives. These two quantities are related but they are not the same.

We shall refer to the pair of a mechanism  $M$  and the distribution  $\mathcal{D}$  as a *sponsored-search environment*. Together, these elements induce distributions over the quality of the advertisement in each slot. We will refer to the assignment of ads to display slots as a ranking, where the first ranked item is assigned to the first slot, etc. For example, if a mechanism produces a ranking over the  $N$  advertisers which is ordered by quality then the distribution of ad quality in the first slot will be the first order statistic of  $N$  draws from  $\mathcal{D}$ , the second slot will be distributed according to the second order statistic, and so forth. We will let the random variable  $Q_i$  denote the quality of the ad shown in the slot  $i$ .

When a user visits the search engine, the search engine displays the ranking produced by  $M(\bar{v}, \bar{q})$  where  $\bar{v}$  is the reported vector of values and  $\bar{q}$  is the reported vector of qualities. We assume that the user starts searching at the top position and proceeds top-down. When a user *views* an ad in position  $i$ , he pays the cost of viewing  $c_v$ , reads the short text of that ad and learns its quality  $q_i$ . When a user *clicks* on an ad in position  $i$ , he pays cost of clicking  $c_c$ , visits the site associated with that ad and has his search need satisfied with probability  $q_i$ . We assume users can only click upon ads that they have already viewed and that they view ads in order (i.e. they cannot view the 3rd ad if they have not viewed the second)<sup>3</sup>. If his search need is satisfied, he receives a (normalized) reward of 1. We assume the user has knowledge of the sponsored-search environment, i.e. the mechanism  $M$  and distribution  $\mathcal{D}$ , and thus knows the distributions for  $Q_1, Q_2, \dots$ . Furthermore, we assume the user is able to update his prior for  $Q_k$  based on the realized qualities of the first  $k - 1$  ads.

This model presents two analytic challenges: (1) determining the distributions of ad qualities for each slot for the given sponsored-search environment and (2) solving for the optimal user search strategy for the given quality distributions in each slot. We treat these two problems separately for the remainder of this paper. In the next section, we present the abstraction for the user search behavior problem but first we present an example to highlight some of the aspects of a search policy.

*Example.* Here we briefly mention an example in which a user should not click on an ad immediately after it is viewed. Suppose that there are two ads, the user viewed the first ad already, and learned  $q_1 = 0.3$ . Conditioned upon this information, the user knows that the quality of the second ad  $Q_2$  is distributed uniform  $[0, 1]$ . The cost of viewing is  $c_v = 0.1$  and the cost of clicking is  $c_c = 0.2$ . The options of the user are then:

<sup>2</sup>It is important to note that we don't study whether these general ranking rules could arise in equilibrium

<sup>3</sup>This assumption is not without loss; there are some environments where this leads to sub-optimal search strategies. Characterizing environments where this restriction affects the optimal policy is an interesting direction for future work

- **click-first**: this incurs a cost of 0.2 with an expected gain of 0.3. In the ad does not satisfy the user, then the user will view the next ad and click if its quality is above .2, giving an expected quality of .6. Thus the expected utility of this strategy is  $.3 - .2 + (.7)(-.1 + .8(.6 - .2)) = .254$ .
- **view-first**: this first incurs a cost of 0.1. The second ad has a quality  $Q_2$  uniformly distributed in  $[0, 1]$ . If this quality is more than 0.3, the user will click on this ad first, and on a failure, will also click on the first ad. This results in a utility of  $q_2 - c_c + (1 - q_2)(q_1 - c_c)$ . The expectation of  $Q_2$  in this range is .65, so this expression is  $.65 - .2 + (1 - .65)(.3 - .2) = 0.485$ . If  $q_2 \in [0.2, 0.3]$ , then the user will click on the first ad, and then, if not successful, click on ad 2. In this case, one can calculate in a similar way that the expected utility is 0.135. Finally, when  $q_2 < 0.2$ , the user will only click on the first ad, resulting in an expected utility of  $q_1 - c_c = .1$ . The overall utility of this strategy is  $-.1 + 0.7 \times 0.485 + 0.1 \times 0.135 + .2 \times .1 = 0.273$ .

Since the expected reward of viewing first is higher, we should expect the user to view the next ad rather than clicking on the first. Then it will click on the second ad first when its ad quality is above that of the first ad, which happens with probability .7 and thus we would actually find a CTR for the first ad that is actually lower than the CTR for the second ad, if we reached this point in the process.

### 3.2. Coin Flipping Problem

In this section, we present an abstract stochastic problem that we will use to generalize the search behavior of users. We feel this problem potentially applies to a few problems. We use it to solve for the optimal, or near optimal, search behavior for users in the sponsored-search context.

**DEFINITION 3.1 (THE COIN FLIPPING PROBLEM).** *Let  $Q_1, Q_2, \dots$  be a countable stream of weighted coins, where  $Q_i$  is the random variable representing the quality (success probability) of the  $i$ th coin in the stream. The cost of drawing a new coin from the stream is  $c_v$  and the cost of flipping a coin is  $c_c$ . The user knows the distributions for  $Q_1, Q_2, \dots$  (and can update his priors based on realized values) but only knows the realized quality of coin  $i$  after drawing it from the stream. At each point, the user has the choice of actions to either flip any coin in his possession, draw the next coin from the stream, or quit. Note, the user is not allowed to skip any coin in the stream, so he must always draw the very next coin if he chooses to draw a new coin. If the coin flip succeeds, he receives a payoff of 1 and quits. If it fails, he loses that coin and continues to the next stage. The user seeks to maximize his expected reward minus costs from this process.*

To analyze this problem, we will frame it as discrete-time stochastic control problem; our actions at time  $t$  will determine both a reward and how the system state evolves from time  $t$  to  $t + 1$ .

The current state  $S = (O, C, k)$  of this process is defined by a set of observed qualities  $O$ , the set of coins in the user's possession  $C$ , and the index of the next coin in the stream  $k$ . We'll denote the set of all states by  $\mathcal{S}$  and write state  $S^0 = (\emptyset, \emptyset, 1)$  to be the starting state of the system. The set of actions consists of  $\{\text{flip any } q \in C, \text{draw, quit}\}$  and is denoted by  $\mathcal{A}$ . The choice of an action determines both the immediate reward received (or cost incurred) and determines the transition probabilities to the next state of the system. The rewards and transitions for each action in the Coin Flipping Problem are as follows.

- (1) Draw next coin. Pay cost  $c_v$ . With probability  $P(Q_k = q_k)$ , transition to state  $S' = (O \cup q_k, C \cup q_k, k + 1)$ . In an abuse of notation, we will abbreviate this new state by writing  $S \cup q_k$ .

- (2) Flip a coin  $q \in C$ . Pay cost  $c_c$ . With probability  $q$ , transfer to the winning state, receive a reward of 1, and end the process. With probability  $1 - q$  transition to state  $S' = (O, C \setminus q, k)$ . We will abbreviate this state by writing  $S \setminus q$ .
- (3) Quit. The user receives a final reward of 0 and the process ends.

A *policy*  $\pi : S \rightarrow \mathcal{A}$  is a function specifying which action  $a$  to take in state  $S$ . The goal of the Coin Flipping problem is to find a policy which maximizes the user's expected return.

LEMMA 3.2 (FLIPPING REWARD). *When taken an action of flip, we pay cost  $c_c$  and with probability  $q$ , transition to the winning state and receive a reward of 1. We transition to state  $S \setminus q$  with probability  $1 - q$ . This is equivalent in expectation to receiving a reward of  $q - c_c$  and transition to state  $S \setminus q$  with probability  $1 - q$ .*

DEFINITION 3.3 (VALUE FUNCTION). *The value function  $V_\pi(S)$  for policy  $\pi$  is a mapping of the expected reward the user obtains when starting in state  $S$  and following policy  $\pi$  for the duration of this process. The expected value for state  $S$  is equal to the reward in the current state plus the future earning from following  $\pi$ . For the Coin Flipping Problem, it has the following form*

$$V_\pi(S) = \begin{cases} q - c_c + (1 - q)V_\pi(S \setminus q) & \text{if } \pi(S) = \text{flip coin } q \\ -c_v + E[V_\pi(S \cup Q_{k+1})] & \text{if } \pi(S) = \text{draw next coin} \\ 0 & \text{if } \pi(S) = \text{quit} \end{cases}$$

where the expectation in the second case is taken over the realized value of coin  $Q_{k+1}$ .

Thus the goal of the Coin Flipping Problem is to find a policy  $\pi$  such that solves

$$\max_{\pi \in \Pi} V_\pi(S^0)$$

where  $\Pi$  is the set of all policies.

The example from the last section demonstrates an important observation; if we set  $c_c = c_v = 0$ , then both strategies would have given the same value because  $q_1 + (1 - q_1)q_2 = q_1 + q_2 - q_1q_2 = q_2 + (1 - q_2)q_1$ . Then it must be that one strategy is better because it minimizes expected costs. In terms of drawing costs, the click-first strategy is superior because we only expect to incur a drawing cost  $(1 - q_1)c_v$  while the view-first strategy incurs a drawing cost of  $c_v$ . However, the draw-first strategy incurs a lower cost of flipping but the reason is more subtle. As we will show, flipping coins in order from highest to lowest quality minimizes the cost of flipping. Since the quality  $q_1$  was relatively low, we expected the  $Q_2 > q_1$  and the gain of flipping coins in the sorted order offsets the higher flipping costs.

Before moving on, we prove a useful lemma showing that the greedy ordering of coin flips minimizes the cost of flipping.

LEMMA 3.4 (GREEDY ORDERING). *Consider the following problem: Given a finite set of coins  $C$ , find the order flipping order of coins to maximize value. Then the sorted order from highest to lowest quality solves this problem.*

PROOF. WLOG, assume  $q_1 \geq q_2 \dots \geq q_n$ . Consider any arbitrary ordering and let  $a_1$  be the quality of the first coin flipped,  $a_2$  be the quality of the second coin flipped, etc. Consider the expected value for flipping in this order

$$a_1 - c_c + (1 - a_1)(a_2 - c_c) + (1 - a_1)(1 - a_2)(a_3 \dots) = \sum_{i=0}^N \left[ \prod_{j=0}^{i-1} (1 - a_j) \right] (a_i - c_c)$$

First we split up the terms to get

$$\sum_{i=0}^N [\prod_{j=0}^{i-1} (1 - a_j)] a_i - \sum_{i=0}^N [\prod_{j=0}^{i-1} (1 - a_j)] c_c$$

First, we focus on the positive term above. As we noted in the example,  $a_i + (1 - a_i)a_{i+1} = a_{i+1} + (1 - a_{i+1})a_i$ , so we can reverse the ordering of any two elements and not affect the value. So any ordering achieves the same value for the first term in above equation. Next, to maximize the overall value, we want to minimize the value of the second term. The  $(1 - a_1)$  appears in the last  $N - 1$  terms, the  $(1 - a_2)$  appears in the last  $N - 2$  terms, etc. Thus we would should first minimize the term  $(1 - a_1)$ , then minimize  $(1 - a_2)$ , etc. This minimization is achieved by the greedy ordering, i.e.  $a_i = q_i$ . Thus the greedy ordering maximizes the value of this expression.  $\square$

Intuitively, this lemma should imply that the optimal action at any state will either be draw the next coin or flip the highest quality coin in the user's possession. We do not show this implication holds for all optimal policies but we will show it to be true for the policies that we are interested in.

#### 4. APPLICATIONS OF MODEL

In this section, we present two sponsored-search environments, show their transformation to the Coin Flipping problem and solve for the optimal policies. Let  $\mathcal{D} = F \times G$  be a product distribution for (quality,value) pairs. We can consider a mechanisms which ranks ads by  $v^r q^{1-r}$  where  $r$  is some chosen weighting factor between the two quantities. Setting  $r = 0$  induces a sorted ranking of ads by quality. Setting  $r = 1$  ranks completely by quality and induces a uniformly random ranking of ad qualities; the distribution over ad quality is the same for each slot. We show that a simple *threshold* strategy is an optimal search policy for this setting.

We are motivated to study these two environments because the they represent the extremes of rankings for this family of ranking rules. For certain distributions, we show there's a small constant additive loss in user value.

##### 4.1. Rank by Quality

First, we will present what we call the 'rank by quality' environment, when  $r = 0$ . The mechanism  $M$  ranks advertisers completely on the basis of quality; the highest quality ad is assigned to the first slot, second highest to the second slot, etc. With  $N$  advertisers participating,  $Q_i$  is the  $i$ th order statistic of  $N$  independent draws from  $F$ . We denote this value by  $F^{i:N}$ .

When transforming to the Coin Flipping problem, we simply have that  $Q_i = F^{i:N}$ . The optimal policy here is quite simple: draw the next coin and flip it immediately until either: (A) the realized quality of the coin  $q_i < c_c$  or (B) the expected quality of the next coin  $Q_{i+1}$  (conditioned upon the realized qualities so far) falls below  $c_c + c_v$ . When  $N$  is large, the first order statistic is very close to 1. Then with high probability, the user will only need to draw and flip a single coin, yielding a user value of  $1 - (c_c + c_v)$ .

##### 4.2. Rank by Value

Now we present the 'rank by value' environment, when  $r = 1$ . The mechanism  $M$  ranks purely on value; i.e. the highest valued advertiser is assigned to the first slot and so forth. Since ad quality and value are independent and  $M$  ranks on the basis of value, the quality of the ad in slot  $i$  is a simply an independent draw from  $F$ .

We study a Coin Flipping problem where each coin follows the distribution, i.e.  $Q_i = Q$  for all  $i$ , and  $Q_i$  and  $Q_j$  are independent for all pairs  $i, j$ . We assume that there are an infinite number of coins in the stream.

After these assumptions, the current state need only be defined by the set of coins currently held by the users; the set of observations is not necessary because each coin is independent of the qualities of all other coins and index is unnecessary because there is always an infinite number of coins remaining at any state. We take advantage of this reduced state space by formulating this problem as a simple Markov Decision Process and use standard MDP tools to show that a threshold policy is optimal for this setting.

## 5. MARKOV DECISION PROCESSES OVERVIEW

Readers familiar with Markov Decision Processes should feel free to skip this section.

We study this problem by formulating it as a Markov Decision Process (MDP), a framework used to solve discrete-time stochastic control problems. At the beginning of each a period, the particular state of the system is observed and then an action must be taken. Based on the current state and the chosen action, a reward is earned (this reward can be negative) and the probability distribution over the next period's state is determined. Formally, an MDP is defined by a set of states  $\mathcal{S}$ , a set of actions  $\mathcal{A}$ , a reward function  $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ , and a transition function  $P : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$ . When action  $a \in \mathcal{A}$  is taken in state  $S \in \mathcal{S}$ , you receive reward  $R(S, a)$  and the process transitions to state  $S'$  with probability  $P(S, S', a)$ .

The goal is to find a policy  $\pi : \mathcal{S} \rightarrow \mathcal{A}$ , which is a mapping of states to actions taken in that state, which maximizes your expected return. We denote the expected return you would receive if you were to start in state  $S$  and follow policy  $\pi$  throughout this process as  $V_\pi(S)$ .

**DEFINITION 5.1.** *For a given policy  $\pi$ , the value function  $V_\pi(S)$  denotes the expected return received by starting in state  $S$  and following policy  $\pi$ .  $V_\pi(S)$  has the following form*

$$V_\pi(S) = R(S, \pi(S)) + \sum_{S' \in \mathcal{S}} P(S, S', \pi(S)) V_\pi(S')$$

This above form for the value function of  $\pi$  has a simple interpretation; it is the reward earned in the current state plus the expected future reward you receive by following policy  $\pi$ .

The Optimality Equation, denoted  $V(S)$ , represents the maximum reward achievable when starting in state  $S$  and is defined by

$$V(S) = \sup_{\pi} V_\pi(S)$$

Similarly, we can say a policy  $\pi^*$  is optimal when

$$V(S) = V_{\pi^*}(S) \quad \forall S \in \mathcal{S}$$

The following theorem is a classic result of Markov Decision Processes.

**THEOREM 5.1.** *The optimal value function  $V(S)$  represents the optimal reward one can earn when starting in state  $S$ . It has the following form*

$$V(S) = \max_a [R(S, a) + \sum_{S' \in \mathcal{S}} P(S, S', a) V(S')]$$

**DEFINITION 5.2.** *For a given policy  $\pi$ , the value function  $V_\pi(S)$  is said to satisfy the optimality equation if*

$$V_\pi(S) = \max_a [R(S, a) + \sum_{S' \in \mathcal{S}} P(S, S', \pi(S)) V_\pi(S')]$$

**THEOREM 5.2** (ROSS [1983]). *If  $V_\pi(S) \geq 0$  for all states  $S$  and  $V_\pi(S)$  satisfies the optimality equation for all states, then  $\pi$  is an optimal policy.*

The above theorem suggests a way to prove that a policy is optimal; guess a candidate policy, derive its value function, and check that the value function satisfies the optimality equation. We use this approach in the next section to show that a threshold policy is optimal for the ‘rank by value’ setting.

## 6. RANK BY VALUE AS AN MDP

In section 2, we described the state space, action, rewards, transition probabilities, and value function for the Coin Flipping Problem. The value function for policy  $\pi$  had the form

$$V_\pi(S) = \begin{cases} q - c_c + (1 - q)V_\pi(S \setminus q) & \text{if } \pi(S) = \text{flip coin } q \\ -c_v + E[V_\pi(S \cup Q_{k+1})] & \text{if } \pi(S) = \text{draw next coin} \\ 0 & \text{if } \pi(S) = \text{quit} \end{cases}$$

Then the optimal value function  $V(S)$  for the Coin Flipping problem has the form

$$V(S) = \max \begin{cases} q_i - c_c + (1 - q_i)V(S \setminus q_i) & \text{coin } q_i \text{ is flipped} \\ -c_v + E_Q[V(S \cup Q)] & \text{new coin drawn} \\ 0 & \text{quit} \end{cases}$$

Thus a policy  $\pi$  satisfies the optimality equation if the following holds for all states  $S \in \mathcal{S}$ .

$$V_\pi(S) = \max \begin{cases} q - c_c + (1 - q)V_\pi(S \setminus q) \\ -c_v + E[V_\pi(S \cup Q_{k+1})] \\ 0 \end{cases}$$

We now show that a threshold policy, with a particular threshold, constitutes an optimal policy for this problem.

We first derive the value function for the threshold policy and show that the reward function satisfies the optimality equation. Theorem 5 allows us to conclude that the threshold policy is optimal.

**DEFINITION 6.1** (THRESHOLD POLICY). *Let  $S$  be the current state and let  $q^1$  be the maximum quality coin in  $S$ . Then the threshold policy with threshold  $t$ , denoted by  $\pi^t(\cdot)$ , is defined as follows*

$$\pi^t(S) = \begin{cases} \text{flip } q^1 & \text{if } q^1 \geq t \\ \text{draw} & \text{if } q^1 < t \end{cases}$$

Since this process begins with  $S = \emptyset$ , the threshold policy immediately flips any coin with quality  $q \geq t$  and never flips any coin  $q < t$ , implying we can discard any  $q < t$ . After we pay  $c_v$  and draw the first coin, we flip it if it is above the threshold and thus end up with  $S = \emptyset$  again. If it is below the threshold, we discard it and again end up with  $S = \emptyset$ .

$$V_{\pi^t}(\emptyset) = -c_v + P(Q \geq t)E[Q - c_c + (1 - Q)V_{\pi^t}(\emptyset)|Q \geq t] + (1 - P(Q \geq t))V_{\pi^t}(\emptyset)$$

Rearranging this term yields

$$c_v = P(Q \geq t)E[(1 - V_{\pi^t}(\emptyset))Q - c_c|Q \geq t]$$

$$V_{\pi^t}(\emptyset) = 1 - \frac{P(Q > t)c_c + c_v}{E[Q|Q > t]P(Q > t)}$$



Maximizing this term over the choice of threshold  $t$  would yield the policy which is optimal amongst the class of threshold policies. We prove below that it is optimal amongst the class of all policies.

DEFINITION 6.2. *Let  $V_0$  denote the value that satisfies*

$$E_Q[\max\{0, (1 - V_0)Q - c_c\}] = c_v$$

Later we show that  $V_0$  and is the expected reward the user can earn from this process by following an optimal policy. For now, we will use  $V_0$  to define a threshold value.

LEMMA 6.1. *Let  $t^* = \frac{c_c}{1-V_0}$ . Then the threshold policy with threshold  $t^*$  satisfies the optimality equation. That is*

$$V_{\pi^{t^*}}(S) = \max \begin{cases} q^1 - c_c + (1 - q^1)V_{\pi^{t^*}}(S \setminus q^1) \\ -c_v + E_Q[V_{\pi^{t^*}}(S \cup Q)] \\ 0 \end{cases}$$

Where the first term in the max is taken over all coins  $q_i \in S$ . Further  $V_{\pi^{t^*}}(\emptyset) = V_0$ .

PROOF. See appendix.  $\square$

THEOREM 1. *The threshold policy with threshold  $t^*$  is an optimal policy for the ‘rank by value’ setting for the Coin Flipping Problem.*

PROOF. Corollary of previous lemma and Theorem 5.  $\square$

## 7. COMPARISON AND DISCUSSION

With these tools in hand, we can discuss the differences in social welfare between these two settings. In this context, social welfare is the sum of the user’s expected value and the sum of the expected value for each advertiser. Unfortunately, computing a closed form for  $V_0$  is not so simple and thus it is hard to establish a general comparison of the ‘rank by value’ and ‘rank by quality’ mechanism for an arbitrary product distribution  $\mathcal{D}$ . For the purposes of discussion, we evaluate the social welfare of the two mechanisms with an infinite number of advertisers and  $\mathcal{D} = F \times G$ ; the distribution for qualities  $F$  is uniform  $[0, 1]$  and the distribution for values  $G$  is uniform  $[0, 1]$ .

### 7.1. Uniform $[0,1]$

LEMMA 7.1. *The value that the user receives from ‘rank by value’*

$$V_0 = 1 - (c_c + c_v + \sqrt{2c_v c_c + c_v^2})$$

With an optimal threshold of

$$t^* = \frac{c_c}{c_c + c_v + \sqrt{2c_v c_c + c_v^2}}$$

The expected value from the set of advertisers is  $\frac{2}{1+t^*}$  yielding a total expected social welfare of

$$SW_{value} = \frac{2}{1+t^*} + 1 - \frac{(1-t^*)c_c + c_v}{\frac{1}{2}(1-t^{*2})}$$

PROOF. The derivation of the user value for this setting comes from plugging in the uniform distribution to equations at the end of 6. The details of this algebra are deferred to the appendix.

Value is generated for an advertiser whenever a user visits clicks through to their site. Users will continue clicking on ads above the threshold until their need is satisfied; each

ad above the threshold has probability  $E[Q|Q \geq t^*] = \frac{1+t^*}{2}$  of satisfying the user. In expectation, users will click on  $\frac{1}{E[Q|Q \geq t^*]} = \frac{2}{1+t^*}$  ads. Similarly, users will only have to view the first  $\frac{1}{P(Q \geq t^*)E[Q|Q \geq t^*]} = \frac{1}{1-t^{*2}}$ , which is a bounded number of ads. Any ad that is clicked upon must be part of the first  $\frac{1}{1-t^{*2}}$  ads; since ads are ordered by quality and there are an infinite number of ads, all of these advertisers will have a value of 1 for a click. Thus the expected welfare generated from the advertisers is  $\frac{2}{1+t^*}$ .  $\square$

LEMMA 7.2. *The value that the user receives from ‘rank by quality’ is*

$$1 - (c_c + c_v)$$

*The expected value for the set of advertisers is  $\frac{1}{2}$ . This yields an expected welfare of*

$$SW_{quality} = \frac{1}{2} + 1 - (c_c + c_v)$$

PROOF. Ads are sorted by quality and thus the first ad will have a quality of 1. This implies that the user will view and click on a single ad and then stop (because his need is satisfied by that ad with probability 1). Qualities and values are independent, so that first ad will have an expected value of a single draw from  $G$ , which is  $\frac{1}{2}$ .  $\square$

Then the ‘rank by value’ model has a higher social welfare when the following is true

$$\frac{2}{1+t^*} + 1 - \frac{(1-t^*)c_c + c_v}{\frac{1}{2}(1-t^{*2})} > \frac{1}{2} + 1 - (c_c + c_v)$$

We can evaluate this expression numerically and we observe there are settings for which the above inequality hold and other settings in which the inequality does not hold. We provide some intuition here for why this is so: because the loss in user value is not too large, the ‘rank by value’ model has a higher social welfare because the user will be clicking on more than just one ad and generating more value for the advertisers. Furthermore, the ads that the user clicks in the ‘rank by quality’ model have higher values because the ads are ranked by value. In the ‘rank by quality’ model, the user only clicks on a single ad which has lower expected value than any of the ads clicked upon in the ‘rank by value’ model. However, when the user costs are high enough, the inefficiency of the user search offsets any gains to advertisers and the ‘rank by quality’ model has a better welfare. It is likely that neither of these ranking rules are socially optimal for the given  $\mathcal{D}$ ; determining the socially optimal ranking rule for various environments is an interesting direction for future work.

## 8. DISCUSSION AND FUTURE RESEARCH

We presented here general framework for studying the optimal user search behavior for a given sponsored-search environment. By incorporating users’ values for satisfying their search needs and their costs associated with searching, we are able to extend the discussion of social welfare in sponsored-search to include consumers. We evaluated the optimal search and the social welfare for a very specific sponsored-search environments. However our model has two shortcomings: (1) this general framework does not simplify the analysis of more complex environments and (2) the optimal policies may be quite complicated for more complex environments, so requiring optimal search seems like a strong assumption. Our analysis of the ‘rank by value’ rule required an infinite number of ads; addressing the effect of a smaller number of ads on the value of the search policy and the social welfare is another important direction for this work to be practically relevant.

In the course of performing this work, we have seen evidence that threshold policies (with the right threshold) are quite reasonable for settings where the induced ranking of qualities is somewhat sorted by quality in expectation. To formalize this notion more, we have the following conjecture

**CONJECTURE 1 (PROPHET INEQUALITY).** *Assume that  $Q_1, Q_2, \dots$  forms a supermartingale, i.e.  $E[Q_{i+1}|Q_i = q_i, Q_{i-1} = q_{i-1}, \dots] \leq Q_i$ . Then there exists a threshold policy which achieves a constant additive approximation in expectation (over the realization of coin qualities) to a prophet who knew all the coin qualities ahead of time.*

We believe that reasonable assumptions on the sponsored-search environment should produce  $Q_1, Q_2, \dots$  that satisfy the supermartingale property.

**CONJECTURE 1 (SUFFICIENT ENVIRONMENT CONDITIONS).** *If ranking rules are monotone (increasing value or quality should only raise you in the ranking) and advertiser values and qualities are positively correlated, then the induced  $Q_1, Q_2, \dots$  should be a supermartingale.*

## REFERENCES

- AGGARWAL, G., FELDMAN, J., MUTHUKRISHNAN, S., AND PÁL, M. 2008. Sponsored search auctions with markovian users. *Internet and Network Economics*, 621–628.
- ATHEY, S. AND ELLISON, G. 2009. Position auctions with consumer search. Tech. rep., National Bureau of Economic Research.
- EDELMAN, B., OSTROVSKY, M., AND SCHWARZ, M. 2005. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. Tech. rep., National Bureau of Economic Research.
- GOMES, R., IMMORLICA, N., AND MARKAKIS, E. 2009. Externalities in keyword auctions: An empirical and theoretical assessment. *Internet and Network Economics*, 172–183.
- KEMPE, D. AND MAHDIAN, M. 2008. A cascade model for externalities in sponsored search. *Internet and Network Economics*, 585–596.
- LAHAIE, S. AND MCAFEE, R. 2011. Efficient ranking in sponsored search. *Internet and Network Economics*, 254–265.
- LAHAIE, S. AND PENNOCK, D. 2007. Revenue analysis of a family of ranking rules for keyword auctions. In *Proceedings of the 8th ACM conference on Electronic commerce*. ACM, 50–56.
- ROSS, S. 1983. *Introduction to stochastic dynamic programming: Probability and mathematical*. Academic Press, Inc.
- VARIAN, H. 2007. Position auctions. *International Journal of Industrial Organization* 25, 6, 1163–1178.

## Appendix

### A. PROOF OF 6.1

**PROOF.** First we prove that  $V_0 = V_{\pi^{t^*}}(\emptyset)$  by showing that  $V_0$  satisfies the recurrence relation if we assume that  $V_0$  is in fact the value for starting with the empty set and using threshold  $t^*$

$$\begin{aligned} V_{\pi^{t^*}}(\emptyset) &= E[V_{\pi^{t^*}}(Q)] - c_v = P(Q \geq t^*)E[Q - c_c + (1 - Q)V_0] + (1 - P(Q \geq t^*))V_0 - c_v \\ &= V_0 + (1 - V_0)P(Q \geq t^*)E[Q - t^* + t^*|Q \geq t^*] - c_cP(Q \geq t^*) - c_v \\ &= V_0 + (1 - V_0)E[\max(Q - t^*, 0)] + ((1 - V_0)t^* - c_c)P(Q \geq t^*) - c_v \\ &= V_0 + E[\max((1 - V_0)Q - c_c, 0)] - c_v = V_0 \end{aligned}$$

Note  $E[\max((1 - V_0)Q - c_c, 0)] = c_v$  by definition. Using this, we now show that the  $t^*$  threshold policy satisfies the optimality equation.

Let the current state be  $S = \{q_1, q_2, q_3, \dots\}$  and relabel such that  $q_1 \geq q_2 \geq \dots$ . Let  $q_1 \geq q_2 \geq \dots \geq q_x \geq t^*$  and  $q_i < t^*$  for  $i > x$ . We must prove the following

$$V_{\pi^{t^*}}(S) = \max \begin{cases} q_1 - c_c + (1 - q_1)V_{\pi^{t^*}}(S \setminus q_1) \\ -c_v + E_Q[V_{\pi^{t^*}}(S \cup Q)] \\ 0 \end{cases}$$

First we prove that flipping  $q_1$  is better than flipping any other coin in  $S$ . If we flip any coin  $q_i \geq t^*$  (the case where  $q_i < t^*$  is obviously not optimal because it results in an expected loss), the next state we move to is  $S \setminus q_i$ . Following policy  $\pi^{t^*}$  at state  $S \setminus q_i$  will result in flipping all the coins above  $t^*$ . Then regardless of what coin when flip at set  $S$ , we will flip all of the coins with quality greater than  $t^*$ , so the set of flipped coins will be the same. The greedy lemma from section 3 implies the value from flipping a set of coins is maximized with the sorted order.

**Case 1:  $q_1 \geq t^*$ .** Next we need to prove that when  $q_1 \geq t^*$ , it is better to flip  $q_1$  than to draw the next coin. That is we need to show the following inequality

$$q_1 - c_c + (1 - q_1)V_{\pi^{t^*}}(S \setminus q_1) \geq E_Q[V_{\pi^{t^*}}(S \cup Q)] - c_v \quad (1)$$

Focus on the LHS of the above: If the newly drawn coin is greater than  $q_1$ , then we will flip on it first and then flip on  $q_1$ . This yields

$$\begin{aligned} LHS &= P(Q \geq q_1)E[Q - c_c + (1 - Q)(q_1 - c_c + (1 - q_1)V_{\pi^{t^*}}(S \setminus q_1)) | Q > q_1] \\ &\quad + P(Q < q_1)E[q_1 - c_c + (1 - q_1)V_{\pi^{t^*}}(\{S \cup Q\} \setminus q_1) | Q < q_1] \end{aligned}$$

We can use the fact that  $(1 - Q)V_{\pi^{t^*}}(S \setminus q_1) = -Q + c_c + V_{\pi^{t^*}}(\{S \cup Q\} \setminus q_1)$  when  $Q > q_1$  to rewrite the above as.

$$\begin{aligned} &= E[q_1 - c_c + (1 - q_1)V_{\pi^{t^*}}(\{S \cup Q\} \setminus q_1)] + P(Q \geq q_1)E[c_c(Q - q_1) | Q \geq q_1] \\ &= q_1 - c_c + (1 - q_1)E[V_{\pi^{t^*}}(\{S \cup Q\} \setminus q_1)] + c_cE[\max(0, Q - q_1)] \end{aligned}$$

So we can write the RHS - LHS of equation 1 as

$$RHS - LHS = (1 - q_1)E[V_{\pi^{t^*}}(S \setminus q_1) - V_{\pi^{t^*}}(\{S \cup Q\} \setminus q_1)] - c_c[\max(0, Q - q_1)] + c_v$$

Consider the  $V_{\pi^{t^*}}(S \setminus q_1) - V_{\pi^{t^*}}(\{S \cup Q\} \setminus q_1)$  term: for any  $Q \leq t^*$ ,  $V_{\pi^{t^*}}(S \setminus q_1) = V_{\pi^{t^*}}(\{S \cup Q\} \setminus q_1)$  and for any  $Q \geq t^*$ , the loss in value only happens when  $Q$  is flipped. Let  $A$  represent the product of  $(1 - q_i)$  for all  $q_i \in S \setminus q_1$  such that  $q_i > Q$ . Then

$$\begin{aligned} V_{\pi^{t^*}}(S \setminus q_1) - V_{\pi^{t^*}}(\{S \cup Q\} \setminus q_1) &= A(Q - c_c - (1 - Q)V_{\pi^{t^*}}(S')) - A(V_{\pi^{t^*}}(S')) \\ &= A(Q - c_c - QV_{\pi^{t^*}}(S')) \end{aligned}$$

where  $S'$  is the state after  $Q$  is flipped. The  $A$  term is at most 1 and  $V_0 \leq V_{\pi^{t^*}}(S)$  for any  $S$ , so we get that  $V_{\pi^{t^*}}(S \setminus q_1) - V_{\pi^{t^*}}(\{S \cup Q\} \setminus q_1) \leq Q - c_c - QV_0$ . Using this, we get

$$\begin{aligned} RHS - LHS &\geq (1 - q_1)P(Q \geq t^*)E[-Q + c_c + QV_0 | Q \geq t^*] - c_cE[\max(0, Q - q_1)] + c_v \\ &= (1 - q_1)(1 - V_0)P(Q \geq t^*)E[-Q + t^* | Q \geq t^*] - c_cE[\max(0, x - q_1)] + c_v \\ &= -(1 - q_1)(1 - V_0)E[\max(0, Q - t^*)] - c_cE[\max(0, Q - q_1)] + c_v \\ &= -(1 - q_1)E[\max(0, (1 - V_0)Q - c_c)] - c_cE[\max(0, Q - q_1)] + c_v \\ &= q_1c_v - c_cE[\max(0, Q - q_1)] \end{aligned}$$

This is an increasing function of  $q_1$  so just consider the case where  $q_1 = t^*$

$$RHS - LHS \geq t^*c_v - c_cE[\max(0, Q - t^*)] = \frac{c_c}{1 - V_0}(c_v - E[\max(0, (1 - V_0)Q - c_c)]) = 0$$

Where the last equality holds because  $V_0$  was defined as the value that solves  $E_Q[\max\{0, (1 - V_0)Q - c_c\}] = c_v$ .

**Case 2:  $q_1 < t^*$**  Consider the case where  $q_1 < t^*$ . Then we need to show the value for flipping is less than the value for drawing. Since  $q_1 < t^*$ , then all coins have quality less than  $t^*$ , so  $V_{\pi^{t^*}}(S \setminus q_1) = V_0$  because we'll never flip any of the other coins. Therefore the value of flipping is

$$q_1 - c_c + (1 - q_1)V_0 < t^* - c_c - (1 - t^*)V_0 = t^*(1 - V_0) - c_c + V_0 = V_0$$

because  $t^* = \frac{c_c}{1 - V_0}$ . So the value of flipping is less than  $V_0$  and now we need to show the value of drawing is at least  $V_0$ . Since the  $q_1 < t^*$ , all other coins are below the threshold, so they will never be flipped. Therefore  $V_{\pi^{t^*}}(S \cup Q) = V_{\pi^{t^*}}(Q)$ . Then the value of drawing is

$$V_{\pi^{t^*}}(S \cup Q) - c_v = V_{\pi^{t^*}}(Q) - c_v$$

As we showed at the beginning of this proof,  $V_{\pi^{t^*}}(Q) - c_v = V_0$ . Therefore the value of drawing is greater than the value of flipping.

Thus we satisfy the recurrence in both cases and we conclude that  $V_{\pi^{t^*}}$  satisfies the optimality equation.

□