Search Design and Broad Matching*

Kfir Eliaz† and Ran Spiegler‡

January 18, 2015

Abstract

We study the problem of designing a mechanism that allocates firms into search pools, when consumers can only give a noisy signal of what they are looking for. Under random sequential search, we show that it is always possible to design a mechanism that incentivizes firms to behave in a way that maximizes consumer surplus (taking their search costs into account). We then establish a necessary and sufficient condition - in terms of the joint distribution of consumer tastes and signals - for the existence of consumer-optimal mechanisms that also extract the entire surplus of firms. The condition is a simple set of inequalities that involve the relative fractions of consumers who like different products, and the Bhattacharyya coefficient of similarity between their conditional signal distributions. When the condition holds, the twin objectives are implementable by an auction that sells (for a price per-impression) "tickets" to consumers' signals, and is augmented by a "broad match" function that links different signals. Under broad matching, a consumer who submits one signal gets a mixed search pool consisting of firms that submitted a winning bid for other signals.

*This is a radically revised version of a paper that appeared under different titles: "The Market for Keywords" and "Broad Matching and the Market for Search Platforms". We gratefully acknowledge financial support from ISF grant, no. 1153/13. We thank Yair Antler, Yossi Azar, Alex Frug, Stephen Hansen, Sergiu Hart, Sebastien Lahaie, David Pennock, Justin Rao and numerous seminar participants for helpful comments.

†Tel-Aviv University and University of Michigan. Email: kfire@post.tau.ac.il.

‡Tel Aviv University, University College London and CFM. URL: http://www.tau.ac.il/~rani. E-mail: rani@post.tau.ac.il.

1 Introduction

Consider the following basic situation. A consumer has a certain need and looks for a product that will satisfy it. If the consumer were confronted with a particular product, he would be able to identify whether he wants it or not. However, the consumer is unable to provide an exact description ex-ante. Therefore, he submits a "query" that gives a rough idea of what he is looking for. The consumer's coarse description fits several types of products (and many individual products within each category, each provided by a different supplier). If a consumer's query is sent to a central planner, what should the planner do, if he aims to maximize consumer surplus? If he gives the consumer a single item, it may be a poor match. Hence, he may try to provide the consumer with a *set of options* to browse. However, if search is time-consuming, the planner should design the set in order to balance two considerations: maximizing the probability that the consumer will find what he is looking for, and the amount of time it will take him to find it. In other words, the planner's problem is to design the consumer's "search environment".

The consumer's predicament is prevalent in other environments. When an employer approaches a Human Resource agency with an intention to hire a new worker, he is unable to rank all potential employees on the other side of the market; instead, he provides a few desirable characteristics (which are often quite vague) that describe the kind of worker he is looking for. Likewise, when we approach a real-estate agent with an intention to rent an apartment, we are typically able to provide only broad characteristics (location, size, existence of an elevator, etc.). When we look for products on internet platforms, we submit keyword-based queries, which are often vague and imprecise. For instance, suppose that you are looking for a specific piece of instrumental music on YouTube, but you forgot its name and the name of its composer. You would recognize it instantaneously if you heard it, but the only information you can provide in a YouTube search is the genre to which the piece belongs. In all these examples, the consumer submits a signal through a deliberate query. In other cases, the consumer passively conveys information when he accesses the intermediary. For instance, in contemporary online platforms, the "cookies" on users' computers contain information about their navigation history and past transactions, and these may be correlated with their current needs.

We are interested in the design of search environments that maximize consumer surplus. We are motivated by two considerations. First, given the consumers' limitation, it is interesting to speculate what would happen if they collectively designed an

institution that addresses it. Second, real-life search platforms (classified directories, online search engines) try to stand out by building a reputation for helping *consumers* find what they are looking for, because they compete for "single homing" consumers (in contrast, suppliers tend to be "multi-homers" - they offer their merchandise at several platforms). Although we do not explicitly model this competitive force, it is in the background of our analysis.

In our framework, attaining a consumer-optimal allocation of suppliers into search pools requires the provision of appropriate incentives for suppliers to self-select into the right search pools. This follows from our assumption that it is prohibitively costly for a planner to verify whether a firm's reported type fits the signal provided by a consumer. Since consumers' signals are noisy, suppliers have an incentive to present themselves as being relevant to many signals (since there is some chance that the supplier is indeed offering what the consumer is looking for). This may clutter the set of options browsed by the consumer, who may consequently spend too much time searching. We therefore ask, can consumer-optimal search design be *decentralized* via some mechanism that will shape the search environment according to the suppliers' individual incentives? In particular, can this mechanism take the form of a "competitive market", in which firms bid for inclusion in the consumer's search environment?

To address these questions, we study the following simple model. There is a finite set of product types X and a finite set of signals W, $|W| \geq |X|$. Each type of product is offered by a measure one of firms. There is a continuum of consumers - each characterized by a pair (x, w), where $x \in X$ is the only type of product the consumer is willing to consume and $w \in W$ is the signal he is able to convey. The value of a product of any type $y \neq x$ for this consumer is 0, and the value of any x product is 1 with independent probability θ (and 0 with probability $1 - \theta$). The signal w can be interpreted as an active query - e.g., a collection of keywords that the consumer can use to describe what he is looking for. In this sense, the signal captures the consumer's "limited vocabulary". The signal can also represent information that the consumer passively conveys, as in the "cookie" example. The distribution of consumer types is $\mu(x, w)$. When signals are imperfectly informative of consumer preferences, two consumers who submit a given signal w may be interested in different product types.

To model the consumer's search environment, we need to make an assumption about the available *search technology*. As a first step, we focus on the most basic and familiar search technology in the literature, namely *random sequential search*. This means that the central planner is unable to control the order in which the consumer inspects individual products in his search pool (in Section 5, we consider the opposite

extreme, where the planner can determine the order of inspection). Thus, we assume that when a consumer of type (x, w) chooses to search and provides the signal w, he is assigned a "search pool" characterized by its *composition of firm types*: the fraction of x products in the pool is $\lambda(x, w)$. The consumer repeatedly decides whether to take a random independent draw from the pool, at a constant cost per draw $c > 0$. When the consumer finds a product he likes, he transacts with the firm. The value of the transaction for the firm is 1 (we abstract from pricing between the consumer and the firm).

The planner's first-best outcome associates with every signal w a probability distribution $(\lambda^*(x, w))_{x \in X}$ that represents the optimal composition of firm types in the search pool of consumers who send the signal w. This can be viewed as an optimal allocation of the "search pool" associated with each w to a heterogeneous population of firms. In basic textbook object-allocation problems, it is efficient to allocate the object to a single type - the one with the highest willingness to pay. By comparison, in our model the search pool is allocated to a *mixture of firms*; the optimal mixture is monotone in the sense that $\lambda^*(x, w)$ increases with $\mu(x, w)$, but multiple firm types will be present in the pool as long as c is not too high.

We first consider anonymous direct mechanisms, in which firms report their type (an element in X), and consequently they are admitted probabilistically into a collection of search pools and pay a lump-sum transfer to the planner.[1] We then show that if participation constraints are ignored (because, say, the planner has an unlimited budget), the first-best is *always* implementable in symmetric pure-strategy Nash equilibrium by some anonymous direct mechanism. The mechanism we construct applies Vohra's (2011) flow-network representation of IC constraints.

Next, we consider the possibility of maximizing consumer surplus with mechanisms that impose some natural requirements on the transfers that firms pay. One natural requirement is that the transfers extract the firms' entire surplus. Another natural requirement is to use a uniform-price mechanism, namely "simple", constant transfers that do not depend on the firm's report. In particular, is it possible to maximize consumer surplus *without any transfers*?

Remarkably, the *same* necessary and sufficient condition applies to all these restrictions on transfers. For the sake of its presentation here, suppose that c is small enough such that all consumer types are served in the first-best. Let $\mu(x)$ be the proportion of consumers who want a product of type x and let $\mu(w|x)$ be the probability that a consumer submits a signal w, conditional on wanting x. The first-best is implementable

[1] We discuss the monitoring assumptions buried in this restriction in Section 3.

(in symmetric pure-strategy Nash equilibrium) by a mechanism that extracts the entire surplus from firms (or, equivalently, uses a constant transfer) if, and only if, for every pair of product types x, y the following inequality holds:[2]

$$\frac{\mu(x)}{\mu(y)} \cdot \left(\sum_{w} \sqrt{\mu(w \mid x)\mu(w \mid y)} \right)^2 \leq 1 \tag{1}$$

The L.H.S. of this inequality is a product of two terms. The first term measures the "popularity gap" between the two product types x and y in the consumer population. The second term is a measure of the similarity of the conditional signal distributions that characterize the preference types x and y. In fact, it is a conventional measure of similarity of probability distributions, known as the "Bhattacharyya coefficient" (after Bhattacharyya (1943)). This measure is commonly used by recommender systems and other machine-learning applications. In our context, it captures the informativeness of consumers' signals, because the Bhattacharyya coefficient is increasing with Blackwell garbling of the system of conditional probabilities $\mu(w \mid x)$.

Thus, the necessary and sufficient condition captures succinctly two obstacles in attaining the joint objectives of maximizing consumer surplus *and* extracting firms' surplus: *a large popularity gap* between two product types, and *similar conditional signal distributions* of two product types. Although our model is highly stylized and obviously cannot be considered as a faithful description of any concrete setting, we believe that the tension identified by (1) is robust and would survive realistic elaborations.

Is there a natural *indirect* mechanism that attains the planner's twin objectives? We propose an auction format, in which each firm submits a "bid-per-impression" for each signal w independently. The highest bidders for each signal v are allocated (uniformly and at random) a number $b(w \mid v)$ of "tickets" to the search pool associated with w. We refer to b as a "broad-match function", hijacking search-engine jargon. Consumers who submit the signal w search by sequential random draws of "tickets" from their pool; when a ticket is drawn, the firm that holds it pays the planner an amount equal to the bid that entitled it to the ticket. Thus, broad matching allows consumers who send the signal w to encounter firms that submitted a winning bid for some other signal v.

Our notion of broad matching captures an intuitive function of real-life institutions

[2]When c is larger such that some consumer types are not served in the first-best, we need to replace μ with the conditional distribution over served types, and the inequality must hold for all pair of product types that are not excluded by the first-best.

that match agents in two-sided markets, which can be described as "vocabulary expansion". When a prospective buyer looks up "road bikes" in a classified directory, he may see items listed by sellers under "racing bikes", "hybrid bikes" or "fixed gear bikes". Although the buyer and seller used different terms, the directory bridges the gap between their vocabularies in the interests of a potentially good match. Likewise, "organic" online search engines respond to keyword-based queries by taking into account synonyms, typos, and logically or semantically related terms. The broad-match function b plays a similar role in our indirect mechanism. Although the term "broad matching" is derived from search-engine jargon, its role in our model is subtly different from the industry practice, as we explain in Section 4.

We show that when inequality (1) holds, a simple specification of b induces a mechanism that maximizes consumer surplus and fully extracts the firms' surplus in some symmetric pure-strategy Nash equilibrium. The set of signals is partitioned into two: a set of "working" signals with as many signals as there are product types, and a set of "dummy" signals. Each working signal v is assigned a distinct product type $f(v)$. The strength of the broad-match link from each working signal v to *every* $w \in W$ is given by $b(w|v) = \sqrt{\mu(w, x)}$. If v is a dummy signal, then $b(w|v) = 0$ for every w. In the symmetric equilibrium, firms of type x bid only for the working signal v for which $f(v) = x$. The winning bid for each working signal v, which is its "equilibrium price-per-impression", has a simple functional form that incorporates a weighted sum of Bhattacharyya coefficients of similarity between $f(v)$ and all relevant product types. Indeed, as signals become less informative of consumer preferences, the equilibrium prices of signals decrease.

2 The Environment

Let X be a finite set of *product types*, $|X| \geq 2$. For every $x \in X$, there is a measure one of *firms* that offer only that product type (as many units as required, at zero cost). We refer to them as "x-firms". A firm gets a fixed payoff of 1 from any unit it sells (we abstract from product prices). Let W be a finite set of *signals*, where $|W| \geq |X|$. There is a measure one of *consumers*. A consumer type is defined by the pair (x, w), where x is the (only) type of product he is interested in, and w is the signal he can convey regarding his wants. In line with some of the potential applications we have in mind, we will often refer to w as a "query", a "keyword" or "the consumer's vocabulary". Let $\mu \in \Delta(X \times W)$ be the distribution of consumer types in the population. We assume the marginals of μ on X and W have full support. As usual, denote $\mu(x) = \sum_w \mu(x, w)$

and $\mu(\cdot|x) = (\mu(w|x))_{w \in W}$. The latter is referred to as the conditional signal/query distribution that characterizes the preference type x.

When a consumer of type (x, w) consumes a product of type $y \neq x$, he gets a sure payoff of 0. When he consumes a particular product of type x, he gets a payoff of 1 (0) with independent probability θ $(1 - \theta)$.[3] Thus, each consumer is interested in one *type* of product, but the (independent) probability he likes any individual product of this type is θ. The parameter θ captures idiosyncratic heterogeneity among consumers and firms. This dimension of differentiated taste also justifies why many firms offer the same type of product. Products are "inspection goods": when a consumer encounters a particular product, he immediately recognizes the payoff it generates. However, all he can communicate ex-ante about what he is looking for is encapsulated in the signal w. Note that w does *not* represent the consumer's information about his own preferences, but the information he can provide to *others*. Thus, when the consumer inspects a particular product, this does not cause him to revise his beliefs.

Example: Mozart vs. Stravinsky

The following specification illustrates the primitives of our model and serves as a running example throughout the paper. Consumers are interested in classical music. There are two types of products: musical pieces by Mozart and by Stravinsky. Accordingly, let $X = \{moz, str\}$. The set of signals consists of three keywords, "Mozart", "Stravinsky" and "Classical Music", denoted MOZ, STR and CL respectively. A consumer type (moz, MOZ) (respectively, (str, STR)) is interpreted as someone who likes the music of Mozart (Stravinsky) and knows how to describe this taste. In contrast, the type (moz, CL) (respectively, (str, CL)) is interpreted as someone who likes the music of Mozart (Stravinsky) without realizing that this is his favorite composer, and all he can say ex-ante is that he is interested in classical music. While he can identify whether he likes an individual piece of music when he hears it, the inspection never reveals the identity of the piece's composer, and the consumer is unable to revise his query. Assume $\mu(moz, STR) = \mu(str, MOZ) = 0$ - that is, when a consumer can name a composer, then he must like his music.

Consumers search for a product they like (i.e., a product that gives them a payoff of 1) according to the following process. Let $a(x, w) \in \{0, 1\}$ represent the decision of a consumer of type (x, w) whether to enter the search process, where $a(x, w) = 0$ means that the consumer refrains from searching, and $a(x, w) = 1$ means that he engages in search by providing the signal/query w. In the latter case, the consumer is given

[3] The assumption that consumers and firms have the same transaction payoff is merely to save unnecessary notation, and plays no role in our analysis.

access to a set of products. We refer to this set as the *search pool* associated with w, and define it formally as a probability distribution $(\lambda(x,w))_{x\in X}$, where $\lambda(x,w)$ is the fraction of products of type x in the pool. The consumer repeatedly draws independent random samples from this pool (with replacement), where each draw carries a fixed cost $c > 0$. As soon as the consumer finds a product he likes, he transacts with its seller and terminates his search. Denote $\lambda = (\lambda(\cdot,w))_{w\in W}$ and $a = (a(\cdot,w))_{w\in W}$.

We wish to make a few comments regarding the search process. First, in the equilibria of the models we analyze, there will always be infinitely many firms in consumers' search pools. This means that our sampling-with-replacement assumption is w.l.o.g. It also ensure that if $\lambda(x,w) > 0$, a searching consumer will find a product he likes in finite time with probability one. Second, we describe the consumer's behavior as a once-and-for-all decision whether to enter the search pool, rather than as a sequential decision problem. This is a formal simplification that carries no loss of generality for our purposes - see Section 3.

Finally, our assumption of a random sequential search technology means that the search designer is unable to control the order by which consumers inspect products in their search pool. This is a reasonable assumption for search environments in which inspection is done "offline". For instance, when a consumer obtains a list of specialists from a "Yellow Pages" directory, he can verify whether a specialist is a good match only by physically contacting him. The order in which he will examine the specialists will depend on factors such as their availability and physical locations, which are beyond the control of the directory's designer.

Even in the case of online search - where search pools are ordered by the search engine - Athey and Ellison (2011) list a number of reasons why web users may disobey the order in which links appear on their computer screen: advertisers may attract the user's attention away from more prominent links; some links may be slow or broken; and the user may distrust the search engine's suggested order. From this point of view, our random-search assumption fits a *worst-case* analysis for the design of search environments (it is also computationally cheaper than complete ordering of *all* the firms in the consumer's search pool according to some given characteristics). At any rate, in Section 5 we examine the diametrically opposed case, in which the search designer can perfectly control the consumer's order of inspection. The intermediate cases, which are more realistic for contemporary online search, are left for future research.

2.1 The Bhattacharyya Coefficient

A key issue in this paper will be the distinctiveness of the signal distributions that characterize consumer preference groups. In particular, we are interested in measuring the closeness of the conditional query distributions $\mu(\cdot|x)$ and $\mu(\cdot|y)$, for any given x and y. Of course, there are multiple ways to measure similarity between probability distributions. However, one particular measure turns out to be relevant for the present model. For any pair of products $x, y \in X$, define

$$S(x,y) \equiv \left(\sum_{w \in W} \sqrt{\mu(w|x)\mu(w|y)} \right)^2$$

In the statistics literature, $\sqrt{S(x,y)}$ is known as the *Bhattacharyya coefficient* that characterizes the distributions $\mu(\cdot|x)$ and $\mu(\cdot|y)$.[4] From a spatial point of view, this is an appropriate similarity measure, because $\sqrt{S(x,y)}$ is the direction cosine between two unit vectors in $\mathbb{R}^{|W|}$, $(\sqrt{\mu(w|x)})_{w \in W}$ and $(\sqrt{\mu(w|y)})_{w \in W}$. The value of $S(x,y)$ increases as the angle between these two vectors becomes narrower; $S(x,y) = 1$ if and only if $\mu(\cdot|x) = \mu(\cdot|y)$; and $S(x,y) = 0$ if the two vectors are orthogonal. More importantly, $S(x,y)$ is an appropriate similarity measure given our interpretation of $\mu(\cdot|x)$ and $\mu(\cdot|y)$ as conditional signal functions. The stochastic matrix $(\mu(\cdot|x))_{x \in X}$ is an information system in Blackwell's sense. This leads to the following observation. (The proofs of all results in this section are relegated to the appendix.)

Remark 1 *When $(\mu(\cdot|x))_{x \in X}$ is subjected to Blackwell garbling, $S(x,w)$ weakly increases for all x, y.*

Thus, Bhattacharyya similarity is increasing as signals become less informative in Blackwell's sense. This is consistent with the intuition that similarity between conditional query distributions captures consumers' limited ability to describe what they are looking for. In our "classical music" example, the assumptions we imposed on μ means that $S(moz, str) = \mu(CL|moz)\mu(CL|str)$.

[4] See Basu, Shioya and Park (2011) and Theodoris and Koutroumbas (2008). A related concept is the *Hellinger distance* between distributions, given by $H^2(x,y) = 1 - \sqrt{S(x,y)}$. In the concluding section we describe practical applications of these measures in recommender systems, and explain how these are related to our model.

2.2 Consumer-Optimal Search Pools

We consider a central planner whose object of design is the composition of search pools. Therefore, it is apt to call him a "*search designer*". The search designer's objective is to maximize consumer surplus, defined as follows:

$$U(\lambda, a) \equiv \sum_{w} \sum_{x} \mu(x, w) u_{\lambda}(x, w), \tag{2}$$

where

$$u_{\lambda,a}(x, w) = \begin{cases} 0 & if \quad a(x, w) = 0 \\ -\infty & if \quad a(x, w) = 1 \text{ and } \lambda(x, w) = 0 \\ 1 - \frac{c}{\theta \cdot \lambda(x,w)} & if \quad a(x, w) = 1 \text{ and } \lambda(x, w) > 0 \end{cases} \tag{3}$$

is the net surplus of consumer type (x, w) under (λ, a). In particular, when $\lambda(x, w) > 0$, a searching consumer of type (x, w) will eventually find a product he likes, after an expected search time of $1/\theta\lambda(x, w)$. Let $(\lambda^*, a^*) = \arg\max_{\lambda,a} U(\lambda, a)$ be the consumer-optimal outcome, defined by the collection of search pools associated with each signal and the consumers' search decisions.

We characterize (λ^*, a^*) in a sequence of steps. First, it is clear from (3) that $\lambda^*(x, w) > 0$ if and only if $a^*(x, w) = 1$. Therefore, from now on we analyze the problem as if we only need to find λ^*. Note that we can calculate $\lambda^*(\cdot, w)$ *independently for each* w.

Second, observe that if $\lambda^*(x, w) > 0$, then $1 - c/\theta\lambda^*(x, w) > 0$. To see why, imagine that $1 - c/\theta\lambda(x, w) \leq 0$, and suppose that we deviate by removing all x products from the pool associated with w (and accordingly switch to $a(x, w) = 0$). This would weakly increase the net surplus of (x, w) consumers. In addition, it would eliminate the negative search externality that x products in the pool exert on consumers who like other types of products. We conclude that if λ^* maximizes total consumer surplus, then $a^*(x, w)$ is also *individually rational* for every consumer type (x, w).

Third, first-order conditions imply that whenever $\lambda^*(x, w)\lambda^*(y, w) > 0$,

$$\frac{\lambda^*(x, w)}{\lambda^*(y, w)} = \sqrt{\frac{\mu(x, w)}{\mu(y, w)}} \tag{4}$$

Since $\sum_{x \in X} \lambda^*(x, w) = 1$, we obtain that whenever $\lambda^*(x, w) > 0$,

$$\lambda^*(x,w) = \frac{\sqrt{\mu(x,w)}}{\sum_{y|\lambda^*(y,w)>0}\sqrt{\mu(y,w)}} \tag{5}$$

The fourth and last step characterizes the set of product types x for which $\lambda^*(x,w) > 0$. We begin by noting the following property of efficient search pools.

Lemma 1 *If $\lambda^*(x,w) = 0$ and $\mu(y,w) < \mu(x,w)$, then $\lambda^*(y,w) = 0$.*

This lemma has the following implication. For each w, order product types in decreasing order of popularity, and denote $\mu_i = \mu(i,w)$, such that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{|X|}$. (Accordingly, denote $\lambda^*(i,w) = \lambda_i^*$.) Then, for each w, there exists a cutoff type m^* such that $\lambda_i^* > 0$ for $i \leq m^*$ and $\lambda_i^* = 0$ for $i > m^*$. The optimal cutoff type is characterized as follows.

Proposition 1 *The cutoff m^* is the highest $m \in \{1, ..., |X|\}$ for which*

$$\sqrt{\mu_m} > \frac{2c}{\theta - c}\sum_{i=1}^{m-1}\sqrt{\mu_i} \tag{6}$$

We introduce a few useful pieces of notation. First, let

$$K = \sum_{(x,w)|a^*(x,w)=1}\mu(x,w) \tag{7}$$

be the total measure of consumers who search under (λ^*, a^*). Second, let μ^* be the consumer type distribution conditional on not being excluded by a^*. That is,

$$\mu^*(x,w) = \frac{\mu(x,w)}{K} \tag{8}$$

We say that a product type x is excluded by λ^* if $\lambda^*(x,w) = 0$ for all $w \in W$. Let X^* be the set of product types that are not excluded by λ^*. When c is small enough, $X = X^*$ and $K = 1$, such that $\mu^* = \mu$.

In our "classical music" example, $\lambda^*(str, STR) = \lambda^*(moz, MOZ) = 1$, because the signals MOZ and STR are perfectly informative of the consumer's preferences. In addition, when c is small, all CL consumers engage in search (recall that this is also individually rational for them), and the fraction of moz products in the search pool associated with the query CL is

$$\lambda^*(moz, CL) = \frac{\sqrt{\mu(moz,CL)}}{\sqrt{\mu(moz,CL)} + \sqrt{\mu(str,CL)}} \tag{9}$$

3 Mechanism Design

We are interested in situations where only a consumer can verify whether he likes a particular product. Moreover, the only means of verification is personal inspection via sequential search. The search designer cannot monitor the composition of product types in a search pool (or it is prohibitively costly for him to do so). Thus, in order to design search pools with a particular composition, firms need to be incentivized to enter the appropriate search pools via some mechanism.

A mechanism requires each firm to submit a message that belongs to some message space (the firm can also choose not to participate, in which case it earns 0). For every profile of messages, the mechanism allocates each firm (probabilistically) to some subset of the search pools associated with each $w \in W$, and specifies a (possibly negative) transfer that the firm pays to the search designer. In principle, the transfer could depend on consumers' behavior, but as we will soon see, this feature will be redundant for our purposes.

Consumers are not players in the mechanism - they decide, simultaneously with the firms' moves, whether to enter the search pool. Their decision is based on rational expectations: they know the search-pool composition λ that is induced by the mechanism and the firms' equilibrium behavior, and a consumer of type (x, w) chooses to search if and only if $1 - c/\theta\lambda(x, w) > 0$. Note that this decision rule is also *sequentially rational* - i.e., the same inequality would govern the consumer's decision whether to continue searching after any number of unsuccessful draws. This justifies our static treatment of consumer choice.

Given that we are interested in implementing the consumer-optimal outcome, we can impose a simplification that carries no loss of generality. We will take it as given that for every consumer type (x, w), $a(x, w) = 1$ if and only if $\lambda^*(x, w) > 0$. We already saw that this behavior is individually rational for consumers given λ^*. This allows us to focus entirely on firms' behavior.

The assumption that consumers have rational expectations can be objected to on the following grounds: if a consumer is unable to provide an exact description of what he is looking for, can he figure out how long it will take to find it? One justification is that consumers often use the same generic query for many search problems. Imagine that you heard a nice song on the radio, and all you can say about the song is that it is an R&B piece by an unknown female singer. Therefore, when you look it up on TouTube, you may enter the string of keywords "R&B female singer". Previous cases in which you submitted this query enable you to estimate its effectiveness.

We could relax the rational-expectations assumption altogether, and simply assume that a consumer type (x, w) submits the query w automatically, without performing any cost-benefit analysis. From a descriptive point of view, this would be equivalent to the case of a sufficiently small c in our model. For the normative analysis, however, this alternative assumption raises the concern whether minimizing expected search time is a "legitimate" welfare criterion, from a revealed-preference point of view.

In the rest of this section, we focus on *direct* mechanisms - i.e. the firms' message space is X. We also restrict attention to *anonymous* mechanisms, in the sense that the outcome for each firm only depends on its own report and the overall distribution of reports. The solution concept we will use is *symmetric pure-strategy Nash equilibrium.* Symmetry means that all firms of a given type behave identically. The revelation principle applies, hence we will assume that firms report truthfully w.l.o.g. (Symmetry or purity of the participation decision is not w.l.o.g - see our discussion at the end of this section.)

Let the mechanism assign a non-negative transfer to any firm that reports $x \notin X^*$. This specification, coupled with consumers' behavior, ensures that a firm of type $x \notin X^*$ will indeed find it optimal not to participate. This enables us to focus exclusively on the behavior of firm types in X^*, and ignore all other agents. From now on, we will take it for granted that *all firms whose type is in X^* participate and report truthfully.* It remains to check their participation and IC constraints.

Given the strategy profile, each x-firm pays a transfer denoted T_x to the search designer; and for any given set of signals V, a fraction $r_x(V)$ of x-firms enters the collection of search pools associated with V. Anonymity means that given this strategy profile, for any individual firm that reports $\hat{x}$, the probability that it enters the collection of search pools associated with V is $r_{\hat{x}}(V)$, and the transfer it will pay is $T_{\hat{x}}$. We can thus reduce any anonymous direct mechanism to the collection $(r, T) = ((r_x)_{x \in X}, (T_x)_{x \in X})$.

The allocation rule r induces the marginal probability that a firm reporting x will enter the search pool associated with any signal w :

$$q(x, w) \equiv \sum_{V \subseteq W | w \in V} r_x(V) \tag{10}$$

We can now establish a simple condition on the function q in order for the induced collection of search pools to coincide with the first-best:

$$\lambda^*(x, w) = \frac{q(x, w)}{\sum_y q(y, w)} \tag{11}$$

Comment: What can the search designer monitor?
Our formal definition of direct mechanisms specifies lump-sum transfers that are independent of events that unfold inside each search pool. Taken literally, this fits situations in which the search designer can only monitor whether agents *access* search pools. This is the case of a "Yellow Pages" directory, for example. However, the definition also fits situations in which the designer can monitor "*impressions*" - i.e., he can condition the firm's transfer on the *number of draws* it receives in each search pool. Online search engines fall into this category. Since in our model firms are risk-neutral and make single decisions, all they care about is the total number of transactions minus the total payment induced by any message they submit. Therefore, our reduction to lump-sum transfers is a mere simplification that carries no loss of generality - we could formulate the mechanism equivalently in terms of transfers that condition on the number of draws.

Things are different when the designer can monitor *transactions*. This assumption is sensible in environments like real-estate intermediation, where transactions are verifiable and exclusive-dealership arrangements enable real-estate agents to prevent the transacting parties from "cutting the middleman" after they are matched. In our context, this assumption would completely trivialize the search designer's problem. By assumption, all firms earn a payoff of 1 conditional on a transaction, regardless of the types of the transacting parties. Therefore, the designer could simply set a uniform price-per-transaction of 1, and thus implement any outcome he wishes. It follows that the interest in our problem arises only when the designer is unable to monitor transactions.

3.1 Implementing Consumer-Optimal Search Pools

Let us first assume that the search designer is only interested in maximizing consumer welfare. Assume in addition that the designer has no budget constraint, hence the firms' participation constraint can always be satisfied. The designer's problem is to find an anonymous direct mechanism that sustains the consumer-optimal collection of search pools λ^* in some Nash equilibrium in which all firms participate and report truthfully. The problem can be stated as follows: find a collection $(r_x, T_x)_{x\in X}$ that induces λ^* according to (10)-(11), subject to the firms' *incentive compatibility* constraint - for every $x, y \in X^*$,

$$\sum_{V\subseteq W} r_x(V) \sum_{w\in V} \frac{\mu(x,w)a^*(x,w)}{\sum_{V\subseteq W|w\in V} r_x(V)} - T_x \geq \sum_{V\subseteq W} r_y(V) \sum_{w\in V} \frac{\mu(x,w)a^*(x,w)}{\sum_{V\subseteq W|w\in V} r_x(V)} - T_y \quad (12)$$

This inequality represents the condition that an x-firm prefers reporting truthfully to pretending to be a y-firm. To understand this IC constraint, recall that by construction, we are considering strategy profiles in which all firms whose type is in X^* participate and report truthfully. The allocation rule $r_{\hat{x}}$ is a probability distribution over subsets of W, induced by the firm's report $\hat{x}$. Consider a realization V of $r_{\hat{x}}$, and let us calculate the number of transactions that an x-firm makes in the search pool associated with some $w \in V$. The total measure of consumers at the pool who are interested in the firm's type of product is $\mu(x,w)a^*(x,w)$. Each of these consumers eventually finds a product that he likes in the pool, and therefore they are equally shared by all the x-firms in the pool. The total measure of these firms is $\sum_{V \subseteq W | w \in V} r_x(V)$. The total number of transactions that an x-firm thus expects to get in the pool associated with w is $\mu(x,w)a^*(x,w)/\sum_{V \subseteq W | w \in V} r_x(V)$. Summing over all $w \in V$, we obtain the total number of transactions that the firm gets when the realization of $r_{\hat{x}}$ is V. The ex-ante expected number of transactions is then calculated w.r.t. $r_{\hat{x}}$. The firm's net payoff is the expected number of transactions minus the transfer $T_{\hat{x}}$ induced by the firm's report.

Let us introduce a few substitutions that simplify the IC constraint. First, recall that $a^*(x,w) = 1$ if and only if $\lambda^*(x,w) > 0$. Second, plug the definition (10). Finally, we can use the definitions (7)-(8). Then, we can rewrite (12) as follows:

$$\sum_w q(x,w) \cdot \frac{K\mu^*(x,w)}{q(x,w)} - T_x \geq \sum_w q(y,w) \cdot \frac{K\mu^*(x,w)}{q(x,w)} - T_y \tag{13}$$

Recall the notation $\sum_w \mu^*(x,w) = \mu^*(x)$. The L.H.S of (13) is thus

$$K\mu^*(x) - T_x \tag{14}$$

which represents the net equilibrium payoff of x-firms. This expression is intuitive: in equilibrium, all searching consumers who like x products eventually transact, and they are equally shared by a measure 1 of x-firms. Thus, the number of transactions that each x-firm gets is $K\mu^*(x)$.

The following lemma provides a useful characterization of the IC constraints *at the first-best*. Let $S^*(x,y)$ be the Bhattacharyya coefficient for the conditional signal distributions $\mu^*(\cdot|x)$ and $\mu^*(\cdot|y)$.

Lemma 2 *An anonymous direct mechanism defined by (r,T) implements λ^* in sym-*

metric pure-strategy Nash equilibrium if and only if

$$K\left[\mu^*(x) - \sqrt{\mu^*(x)\mu^*(y)S^*(x,y)}\right] \geq T_x - T_y \tag{15}$$

for every distinct $x, y \in X^*$.

Proof. Plug (11) into (13) and obtain yet another way of writing the IC constraint:

$$K\left[\mu^*(x) - \sum_w \frac{\lambda^*(y,w)}{\lambda^*(x,w)}\mu^*(x,w)\right] \geq T_x - T_y \tag{16}$$

By (4), whenever $\lambda^*(x,w)\lambda^*(y,w) > 0$, we have

$$\frac{\lambda^*(y,w)}{\lambda^*(x,w)} = \frac{\sqrt{\mu^*(y,w)}}{\sqrt{\mu^*(x,w)}}$$

This means that (16) is satisfied if and only if

$$\begin{aligned}
\mu^*(x) &\geq \frac{T_x - T_y}{K} + \sum_w \frac{\sqrt{\mu^*(y,w)}}{\sqrt{\mu^*(x,w)}} \cdot \mu^*(x,w) \\
&= \frac{T_x - T_y}{K} + \sum_w \sqrt{\mu^*(y,w)\mu^*(x,w)} \\
&= \frac{T_x - T_y}{K} + \sqrt{\mu^*(x)\mu^*(y)} \sum_w \sqrt{\mu^*(w|x)\mu^*(w|y)} \\
&= \frac{T_x - T_y}{K} + \sqrt{\mu^*(x)\mu^*(y)S^*(x,y)}
\end{aligned}$$

Rearranging, we obtain (15). ■

The next result establishes that the consumer-optimal outcome is always implementable by some anonymous direct mechanism.

Proposition 2 *There is an anonymous direct mechanism that implements* λ^* *in symmetric pure-strategy Nash equilibrium.*

Proof. We prove a stronger result: the first-best is implementable signal-by-signal. That is, we can design independent mechanisms for each w, as if the entire set of signals consisted of w alone. This result is stronger because we continue to assume the same equilibrium behavior by firms (participation and truthful revelation) and the same implemented outcome, but there are more constraints to satisfy in the signal-by-signal

implementation problem (a separate participation and IC constraint for each signal, compared with a single participation and IC constraint in the grand mechanism).

Note that when the relevant set of signals is a singleton, $S^*(x,y)=1$ for every x,y. Consider some $w \in W$ and x,y for which $\lambda^*(x,w)\lambda^*(y,w)>0$. By Lemma , the IC constraint that prevents type x from pretending to be y, denoted $IC(x,y)$, is given by the inequality,

$$K\left[\mu^*(x)-\sqrt{\mu^*(x)\mu^*(y)}\right] \geq T_x - T_y \tag{17}$$

Let $\phi(x,y)$ denote the L.H.S. of (17), and rewrite the constraint $IC(x,y)$ as $\phi(x,y) \geq T_x - T_y$. For any cycle of products $(x_1, x_2, \ldots x_m, x_1)$,

$$\begin{aligned}
\phi(x_1,x_2)+\cdots+\phi(x_m,x_1) &= K\sum_{i=1}^{m}\left(\mu(x_i)-\sqrt{\mu(x_i)\mu(x_{(i+1)\bmod m})}\right) \\
&\geq K\sum_{i=1}^{m}\left(\mu(x_i)-\frac{\mu(x_i)+\mu(x_{(i+1)\bmod m})}{2}\right) \\
&= 0
\end{aligned}$$

Thus, by Rochet (1987), there exist transfers $(T_x)_{x\in X}$ that satisfy $IC(x,y)$ for all $x,y \in X$. We now explicitly derive these transfers, using Vohra's (2011) flow-network technique.

Consider a complete weighted directed graph, whose set of nodes is X^*, and the weight on the link $x \to y$ is $\phi(x,y)$. Add a link from any x to itself, whose weight is $\phi(x,x)=0$. A *path* from x to y is a sequence of nodes that begins with x and ends with y. Define the *length* of a path to be the sum of the weights on the links along the path. Let $\delta(x,y)$ be the distance from x to y, namely the length of the *shortest path* from x to y. Since the sum of weights along any cycle is non-negative, the distance is always well-defined and non-negative, and by definition it satisfies the triangle inequality: for any x,y,z, $\delta(x,z) \leq \delta(x,y)+\delta(y,z)$. Fix some $x^* \in X$. For any $x \in X^*$, define $T_x = \delta(x,x^*) - L$, where $L>0$ is large enough to ensure that all participation constraints hold. By the triangle inequality, $\phi(x,y)+\delta(y,x^*) \geq \delta(x,x^*)$ for any $x,y \in X^*$. This implies that for any pair of distinct products x,y in X^*, $\phi(x,y) \geq T_x - T_y$, hence $IC(x,y)$ is satisfied. ■

Let us apply the explicit mechanism derived in the proof of Proposition 2 to our "classical music" example (given the first-best outcome given at the end of Section 2.2 for small c). The proof established that we can implement the first-best signal by signal, treating each signal as an independent mechanism-design problem. The problem for the signal MOZ is trivial, because all consumers who submit this sig-

nal want moz. Thus, the designer can prescribe $T_{moz} = \varepsilon > 0 = T_{str}$, and if ε is small enough, no firm would have any incentive not to participate or to misreport its type. An analogous argument holds for STR. Let us turn to the mechanism associated with CL. The quantities $\phi(moz, str)$ and $\phi(str, moz)$ defined in the proof are $\mu(moz, CL) - \sqrt{\mu(moz, CL)\mu(str, CL)}$ and $\mu(str, CL) - \sqrt{\mu(moz, CL)\mu(str, CL)}$, respectively. Following the procedure in the proof, we can set $T_{str} = 0$ and $T_{moz} = \phi(moz, str)$, and all participation and IC constraints will hold.

3.2 Full Surplus Extraction and Uniform-Price Mechanisms

In this sub-section we assume that the search designer has additional objectives, apart from implementing consumer-optimality. First, suppose that he wishes to implement λ^* and at the same time extract the firms' entire surplus. Our next result establishes a necessary and sufficient condition for the implementability of these twin objectives.

Proposition 3 *There exists an anonymous direct mechanism that implements λ^* in a symmetric pure-strategy Nash equilibrium in which all firms earn zero payoffs, if and only if*

$$\frac{\mu^*(x)}{\mu^*(y)} S^*(x, y) \leq 1 \tag{18}$$

for every $x, y \in X^*$.

Proof. Let $x, y \in X^*$. In order for x-firms and y-firms to earn zero profits in equilibrium, we must have $K\mu^*(x) - T_x = K\mu^*(y) - T_y = 0$, according to (14). Thus, $IC(x, y)$, as given by (15), is reduced to $\mu^*(y) \geq \sqrt{\mu^*(x)\mu^*(y)S^*(x, y)}$, which is equivalent to (18). ■

Thus, the forces that obstruct implementability of consumer-optimal search pools and full extraction of firms' surplus are large popularity gaps between products and uninformative signals. Consider two preference types x and y. If x is significantly more common than y in the consumer population, and the conditional signal distributions that characterize x and y are relatively similar, $IC(x, y)$ will fail to hold. In our "classical music" example, $IC(moz, str)$ is violated when $\mu(moz)$ and $\mu(CL)$ are large - i.e., when there are many moz fans and many consumers who submit the generic query CL.

A few comments about this result are in order. First, when μ^* is the uniform distribution over X^*, the twin objectives are implementable, because by definition, $S^*(x, y) \leq 1$ for every x, y. Second, when c is large enough, the consumer-optimal

search pools are all homogenous, such that $S^*(x,y) = 0$ for all x, y, and therefore the twin objectives are implementable. Finally, for generic μ and small c, signal-by-signal implementation is impossible - unlike the case where the only objective is maximizing consumer welfare - because when we restrict ourselves to an environment with one signal, $S^*(x,y) = 1$ for all x, y.

Next, suppose that the designer wishes to implement λ^* using a mechanism in which transfers are independent of the firms' reports - i.e., $T_x = T_y$ for all $x, y \in X^*$. In particular, the mechanism may be required to involve no transfers at all. We refer to anonymous direct mechanisms that satisfy this additional requirement as uniform-price mechanisms. It turns out that the condition for implementability of the first-best by a uniform-price mechanism is the same as under the full-surplus-extraction objective.

Proposition 4 *There exists a uniform-price mechanism that implements λ^* in symmetric equilibrium, if and only if (18) holds for every $x, y \in X^*$.*

Proof. Let $x, y \in X^*$. Impose the uniform-price requirement $T_x = T_y$. Then, $IC(y,x)$, as given by (15), is reduced to $\mu^*(y) \geq \sqrt{\mu^*(x)\mu^*(y)S^*(x,y)}$, which is equivalent to (18). (Note that the result follows from examining $IC(y,x)$, whereas in the proof of Proposition (3) it followed from examining $IC(x,y)$.) ■

To appreciate this coincidence, consider an abstract single-agent mechanism-design setting with two agent types, 1 and 2, and two possible outcomes, O_1 and O_2. Suppose that the mechanism designer's objective is to assign the outcome O_k to type k. The types' gross payoff from each outcome is given by the following matrix:

	O_1	O_2
1	d_{11}	d_{12}
2	d_{21}	d_{22}

Let t_k denote the transfer the agent pays when he reports his type to be k. The IC constraints that ensure truthful reporting are:

$$\begin{aligned} d_{11} - t_1 &\geq d_{12} - t_2 \\ d_{22} - t_2 &\geq d_{21} - t_1 \end{aligned}$$

Full surplus extraction implies $t_k = d_{kk}$, and the IC constraints are reduced to $d_{22} \geq d_{12}$ and $d_{11} \geq d_{21}$. Alternatively, uniform prices mean $t_1 = t_2$, and the IC constraints are reduced to $d_{11} \geq d_{21}$ and $d_{22} \geq d_{21}$. Obviously, these systems of inequalities need not

coincide. And indeed, in general settings, the conditions for implementing an outcome with full surplus extraction and with a uniform-price mechanism are different. However, if d_{12} and d_{21} happen to be the same, the two systems do coincide.

This is precisely what happens in our setting. Under λ^*, the number of transactions that an x-firm expects when it pretends to be y is

$$\sum_w q(y,w) \cdot \frac{K\mu^*(x,w)}{q(x,w)} = K \sum_w \frac{\lambda^*(y,w)}{\lambda^*(x,w)} \mu^*(x,w) = K \sum_w \sqrt{\mu^*(x,w)\mu^*(y,w)}$$

This is exactly the same as the number of transactions that a y-firm expects when it pretends to be x:

$$\sum_w q(x,w) \cdot \frac{K\mu^*(y,w)}{q(y,w)} = K \sum_w \frac{\lambda^*(x,w)}{\lambda^*(y,w)} \mu^*(y,w) = K \sum_w \sqrt{\mu^*(x,w)\mu^*(y,w)}$$

This coincidence is a consequence of the first-order conditions that characterize consumer-optimal search pools.

Comment: Asymmetric/Mixed equilibria
Our analysis was based on the restriction that firms play a symmetric pure-strategy Nash equilibrium - which involved, w.l.o.g., truthful reporting. It turns out that if we relax the symmetry or purity of firms' participation decision, the twin objectives of consumer-optimality and full extraction of firms' surplus are always implementable. To see why, note that when firms' surplus is fully extracted, they are indifferent to participation. Suppose that for every x, a fraction $m(x)$ of x-firms participate. Let us continue to use $q(\hat{x}, w)$ to denote the probability that an individual firm reporting $\hat{x}$ enters the search pool associated with w, given the firms' equilibrium behavior. Then, the total measure of x-firms in that pool is no longer $q(x,w)$, but $m(x)q(x,w)$. If we select $m(x) = \mu(x^*)$, we mimic an environment in which μ^* is uniform, where - as we saw - the twin objectives are implementable.

However, this construction relies on an arbitrary selection among a continuum of asymmetric/mixed equilibria, and on taking very literally the bindingness of firms' participation constraint. It is doubtful that this selection can be justified by convincing purification or dynamic stability arguments. Note that we cannot use a similar construction for uniform-price mechanisms, where participation constraints are not binding.

4 Competitive Bidding and Broad Matching

In the previous section, we analyzed the joint implementability of consumer-optimality and full extraction of firms' surplus, focusing on anonymous *direct* mechanisms. Is there a plausible *indirect* mechanism that mimics the optimal direct mechanism when condition (18) holds? In particular, we are interested in mechanisms based on *competitive bidding*, because they are natural benchmarks in object-allocation mechanisms and commonly used by real-life search platforms (notably online search engines).

We introduce a mechanism referred to as a "*broad-match auction*". As before, we take it as given that every consumer type (x, w) plays $a(x, w) = 1$ if and only if $\lambda^*(x, w) > 0$, such that our mechanism focuses on firms' behavior. Each firm simultaneously chooses which signals in W to bid for and how much. Bids are "per impression", according to a rule we explain below. We introduce a *broad-match function* $b : W \times W \to [0, 1]$, which is a weighted directed graph over the set of signals. The broad-match function assigns to each signal $v \in W$ an endowment of "(w, v)-tickets" for every w, the measure of which is denoted $b(w|v)$. Each of these (w, v)-tickets is uniformly assigned among the firms that submitted the highest bid for v.

When a consumer of type (x, w) chooses to search and thus submits the query w, he gets access to a search pool consisting of a measure $\sum_{v \in W} b(w|v)$ of $(w, \cdot)$ tickets, originating from all the signals v via the broad-match link to w. The consumer then sequentially samples at random *tickets* from this pool (with replacement). Each time a consumer draws a (w, v) ticket, he incurs a search cost c and inspects the firm that holds the ticket. For this draw, the firm pays to the search designer the (winning) bid-per-impression it submitted for v. Thus, when ticket that a firm holds is sampled in the search pool associated with w, the "price per impression" it pays depends on the signal v from which the firm's ticket originated.

Note that there are strategy profiles in the game for which the set of highest-bidders for some signal will have a measure zero. In this case, the payoff of these firms, as described above, is ill-defined. All we will need to assume in this case is that if the *net profit per ticket* that a firm holds, thanks to submitting a winning bid for some signal, is negative, then the firm derives a negative payoff from this signal (and would therefore prefer not to bid for it).

This completes the description of the broad-match auction. We discuss the interpretation of our notion of broad matching, and how it is related to the term's usage in the search-engine industry, in Section 4.3. As a benchmark for our analysis, let us consider two extreme specifications of b.

Exact matching

Suppose that b is the degenerate graph - i.e., $b(w|w) = 1$ and $b(w|v) = 0$ for all $w \neq v$. In this case, we have an "exact-match auction": consumers who submit the query w are brought into contact only with firms that submitted a winning bid for w. This reduces our mechanism to a signal-by-signal first-price auction. If the set $\{x \mid \mu^*(x, w) > 0\}$ is a singleton for every w (either because consumers' signals are perfectly informative, or because c is large), the exact-match auction can implement both consumer-optimality and full extraction of firms' surplus in symmetric pure-strategy Nash equilibrium.

The difficulty with exact matching arises precisely when optimal search pools are required to be diverse. Consider the "classical music" example. If we restrict attention to symmetric pure-strategy profiles, $\lambda(\cdot, CL)$ can only take the values $0, \frac{1}{2}, 1$, and therefore it is mechanically unable to generate $\lambda^*(\cdot, CL)$. We later show that the exact-match auction fails to implement the designer's twin objectives, even when we relax symmetry or purity.

Fully broad matching

Suppose that b is the complete graph - i.e., $b(w|v) = 1$ for all w, v. This reduces our mechanism to exact matching, defined for a fictitious environment that consists of one signal w^*, such that the fraction of (x, w^*) consumers is $\mu(x)$. In this case, the mechanism does worse than the above exact-match auction, in the sense that there are μ for which the former fails to implement the twin objectives while the latter succeeds. The reason is that the fully-broad-match function effectively eliminates all the information contained in consumers' signals.

Is there a broad-match function b for which the broad-match auction gives rise to a symmetric pure-strategy Nash equilibrium in which firms earn zero profits and the induced collection of search pools is λ^*? Proposition 3 implies that a necessary condition is that inequality (18) holds for every $x, y \in X^*$. However, it is not obvious a priori that the same condition is also sufficient when we restrict ourselves to broad-match auctions.

4.1 Mozart vs. Stravinsky Revisited

We illustrate the broad-match auction with our "classical music" example. First, we prune the broad-match graph by setting $b(w|CL) = 0$ for all w. This ensures that CL has no value for firms. Accordingly, we construct symmetric pure-strategy Nash equilibria in which no firm bids for CL. Suppose that *moz* (*str*) firms bid exclusively for MOZ (STR). Denote their bids by $p(MOZ)$ and $p(STR)$, respectively. These

are of course the winning bids for the two keywords; we defer their exact specification. Since consumers who submit the query MOZ (STR) are not interested in *str* (*moz*), it makes sense to sever additional links: $b(MOZ|STR) = b(STR|MOZ) = 0$. Next, we normalize $b(MOZ|MOZ) = b(STR|STR) = 1$. The only values of b that are left to be specified are $b(CL|MOZ)$ and $b(CL|STR)$.

Given the firms' assumed behavior and the structure we imposed on b, the search pool associated with MOZ (STR) is homogenous and consists of a measure one of *moz* (*str*) firms, as demanded by λ^*. As to the search pool associated with CL, we have

$$\lambda(moz, CL) = \frac{b(CL|MOZ)}{b(CL|STR) + b(CL|MOZ)}$$

Recall that the consumer-optimal outcome requires

$$\lambda^*(moz, CL) = \frac{\sqrt{\mu(moz, CL)}}{\sqrt{\mu(moz, CL)} + \sqrt{\mu(str, CL)}} \tag{19}$$

Equating the two terms, we obtain the following condition that b must satisfy:

$$\frac{b(CL|STR)}{b(CL|MOZ)} = \frac{\sqrt{\mu(str, CL)}}{\sqrt{\mu(moz, CL)}} \tag{20}$$

Are the firm's individual incentives consistent with this restriction? It turns out that *broad matching creates a new incentive problem that would not exist under exact matching.* The broad-match link $STR \to CL$ means that if an individual *moz* firm deviates by submitting the bid $p(STR)$, it gains access to consumers who submit the query CL. If *moz* is a popular product among this group, the deviating firm can get many transactions from (moz, CL) consumers, and as a result its deviation may be profitable.[5] This is the incentive problem due to broad matching: *sellers of popular product types may bid for keywords they would avoid under exact matching.*

In order to prevent this deviation by a *moz* firm, the probability that a consumer transacts with the firm, thanks to drawing any (w, STR) ticket, cannot exceed $p(STR)$. In other words, the *moz* firm's average "conversion rate" associated with the tickets available at STR must be weakly lower than the equilibrium price-per-impression of STR, which in turn must be weakly lower than the conversion rate of *str* firms. Because (STR, STR) and (CL, STR) tickets are allocated uniformly among all firms that submit the winning bid for STR, *moz* and *str* firms experience the same number of

[5] Since individual firms are non-atomic, the deviating firm will receive a negligible fraction of the available tickets at STR, and therefore the deviation will not affect the composition of any search pool.

impressions thanks to submitting this bid. Therefore, as far as the deviation in question is concerned, we only need to check whether, given the firms' putative equilibrium behavior, $(\cdot, STR)$ tickets do not generate on average a larger number of transactions for an individual *moz* firm than for an individual *str* firm.

Let us first calculate the number of transactions an individual *moz* firm expects if it is among the highest bidders for STR. First, the firm would get access to the CL search pool, where a measure $\mu(moz, CL)$ of consumers demand *moz*. This demand is uniformly distributed over all the tickets in the pool that are held by *moz* firms. Both before and after the deviation, this measure is $1 \cdot b(CL|MOZ)$. Thus, the number of transactions per ticket held by a *moz* firm in the CL search pool is $\mu(moz, CL)/b(CL|MOZ)$. The total number of (CL, STR) tickets is $b(CL|STR)$, uniformly allocated among a measure-one set of firms (both before and after the deviation). The *moz* firm will obtain no transactions in the STR pool. Therefore, the number of transactions that a single *moz* firm expects from paying the winning bid for STR is

$$\frac{b(CL|STR)}{1} \cdot \frac{\mu(moz, CL)}{b(CL|MOZ)}$$

A similar calculation establishes that the number of transactions that a single *str* firm expects from paying the winning bid for STR is

$$\frac{b(STR|STR)}{1} \cdot \frac{\mu(str, STR)}{b(STR|STR)} + \frac{b(CL|STR)}{1} \cdot \frac{\mu(str, CL)}{b(CL|STR)} = \mu(str)$$

There are two terms on the L.H.S because *str* firms will obtain transactions at both CL and STR pools. It follows that the deviation will be unprofitable if and only if

$$\mu(str) \geq \frac{b(CL|STR) \cdot \mu(moz, CL)}{b(CL|MOZ)}$$

Inserting (20), this inequality can be rewritten as

$$\sqrt{\frac{\mu(moz)}{\mu(str)} \cdot \mu(CL|moz) \cdot \mu(CL|str)} \leq 1$$

Since $\mu(MOZ|str) = \mu(STR|moz) = 0$, this is equivalent to

$$\frac{\mu(moz)}{\mu(str)} S(moz, str) \leq 1$$

A similar calculation needs to be carried out to ensure that an individual *str* would not

want to deviate by submitting the bid $p(MOZ)$ for MOZ. The condition that would prevent this deviation is

$$\frac{\mu(str)}{\mu(moz)} S(moz, str) \leq 1$$

But these two inequalities constitute precisely the condition (18)!

We need to check two other types of deviations. First, suppose that a moz firm bids *above* $p(STR)$ for STR. Then, the firm will get all (STR, STR) tickets, thus crowding all str firms out of the pool associated with STR. Because consumers adhere to their equilibrium strategy, (str, STR) consumers will keep searching but they will never find a product they like in their pool. As a result, every (STR, STR) ticket will receive infinitely many impressions, and the average conversion rate experienced by the deviating moz firm would drop to zero, hence the deviation is unprofitable. This is a consideration that restrains overbidding and thus competitive pressures - it is an unusual feature that does not exist in conventional object-allocation problems.

The last type of deviation we need to address is when a moz (str) firm bids above $p(moz)$ ($p(str)$). Note that this deviation does not change the firm's conversion rate, because all $(\cdot, STR)$ and $(\cdot, MOZ)$ tickets are held by the same type of firms before and after the deviation. Therefore, a necessary and sufficient condition for ruling out this deviation is that $p(MOZ)$ and $p(STR)$ are equal to the conversion rates that moz and str firms experience from MOZ and STR, respectively.

Consider the case of STR. The conversion rate is equal to the number of transactions that an individual str firm obtains from STR divided by the number of impressions it obtains from STR. We saw that the former term is $\mu(str)$. The latter term is

$$b(CL|STR) \cdot \left[\frac{\frac{\mu(moz,CL)}{\theta \cdot \lambda^*(moz,CL)} + \frac{\mu(str,CL)}{\theta \cdot \lambda^*(str,CL)}}{b(CL|MOZ) + b(CL|STR)} \right] + b(STR|STR) \cdot \left[\frac{\frac{\mu(str,STR)}{\theta \cdot \lambda^*(str,STR)}}{b(STR|STR)} \right] \quad (21)$$

To understand this expression, note that the impressions that an individual firm gets from the signal STR originate from two search pools: STR and CL. The first term in 21 accounts for the impressions from the CL pool. There are $b(CL|STR)$ "tickets" to this pool, and they are uniformly allocated to a measure one of firms (since in equilibrium only the str firms submit the winning bid for STR). In the CL pool, there are consumers of two type, (moz, CL) and (str, CL). Each consumer of type (moz, CL) ((str, CL)) draws $1/\theta\lambda^*(moz, CL)$ ($1/\theta\lambda^*(str, CL)$) tickets on average until he finds a product he likes. Hence, the total number of draws from all consumers in the CL pool is equal to the numerator of the bracketed expression in the first term of

(21). These draws are equally shared by all the tickets in the CL pool, the measure of which is $b(CL|MOZ) + b(CL|STR)$. It follows that the expected number of draws from the CL pool, received by any firm that submits a winning bid for STR, is given by the first term of 21. By a similar reasoning, the second term of 21 represents the expected number of draws from the STR pool.

Substituting (19)-(20) into (21) yields

$$p(STR) = \frac{\theta}{1 + \sqrt{\frac{\mu(moz,CL)}{\mu(str,CL)}}}$$

A similar calculation would establish that

$$p(MOZ) = \frac{\theta}{1 + \sqrt{\frac{\mu(str,CL)}{\mu(moz,CL)}}}$$

To conclude, in the "classical music" example, our specification of b (which is not uniquely pinned down, because equation (20) only determines the *ratio* between $b(CL|MOZ)$ and $b(CL|STR)$) ensures that the broad-match auction implements the twin objectives of consumer-optimality and full extraction of firms' surplus in symmetric pure-strategy Nash equilibrium, as long as the general necessary and sufficient condition for their implementability is satisfied.

Comment: The limits of exact matching

The need for a suitably designed broad-match function arises not only because of our restriction to symmetric pure-strategy Nash equilibria. To see why, imagine that the designer employs the exact-match auction. Let $m(x)$ denote the measure of x firms in the CL search pool that is induced by some (potentially asymmetric or mixed) Nash equilibrium. Clearly,

$$\frac{m(moz)}{m(str)} = \frac{\lambda(moz, CL)}{\lambda(str, CL)} \tag{22}$$

In order for λ to coincide with λ^*, it must be the case that $m(moz) > 0$ and $m(str) > 0$. Therefore, both moz and str firms submit the winning bid for CL. In order for the firms' surplus to be fully extracted, the winning bid must be equal to the conversion rate they expect at CL. Therefore, they must expect the same number of transactions in the pool associated with CL. This number for an individual x firm is $\mu(x, CL)/m(x)$, because all consumers of type (x, CL) search and eventually find a product they like,

and all x firms equally share this clientele. Plugging (22), we obtain

$$\frac{m(moz)}{m(str)} = \frac{\lambda(moz, CL)}{\lambda(str, CL)} = \frac{\mu(moz, CL)}{\mu(str, CL)} \tag{23}$$

which contradicts the condition for consumer-optimality (19). Thus, for generic μ, the planner's twin objectives cannot be implemented by an exact-match auction.

4.2 A "Canonical" Broad Match Function

We now construct a "canonical" broad-match function b^* that ensures that whenever condition (18) holds, the broad-match auction can implement the twin objectives of consumer-optimality and full extraction of firms' surplus in symmetric pure-strategy Nash equilibrium. Select an arbitrary subset $W^* \subseteq W$ such that $|W^*| = |X^*|$, and let $f : W^* \to X^*$ be an arbitrary one-to-one function. Now define b^* as follows. For every $w \in W$,

$$b^*(w|v) = \begin{cases} 0 & if \quad v \notin W^* \\ \sqrt{\mu^*(f(v), w)} & if \quad v \in W^* \end{cases}$$

The canonical broad-match function has a natural interpretation. The signals outside W^* are "dummy keywords" with no value for bidders, because they grant no tickets to any search pool. The signals in W^* are "working keywords" worth bidding for. The function f assigns meaning to every such query; $f(v)$ is the product type associated with the keyword v. The strength of the broad-match link from v to w increases with the fraction of $(f(v), w)$ types among searching consumers. As we explain below, it is important that $b^*(w|v)$ is proportional to $\sqrt{\mu^*(f(v), w)}$.

Proposition 5 *Consider the following strategy profile. For every $x \in X^*$, all x-firms bid*

$$p(v) = \frac{\theta}{\sum_y \sqrt{\frac{\mu^*(y)}{\mu^*(f(v))} S^*(f(v), y)}}$$

for the signal $v = f^{-1}(x)$, and do not bid for any other signal. If $x \notin X^*$ then x-firms do not bid for any signal. This strategy profile constitutes a Nash equilibrium in the broad-match auction defined by b^*. In this equilibrium, each firm earns zero profits (such that the firms' entire surplus is surrendered to the search designer) and the induced collection of search pools is λ^*. (The proof appears in the appendix.)

Thus, the canonical broad-match function addresses the incentive problem due to broad matching, to the extent possible given Proposition 3. On one hand, allocation of

signals to firms is based on competitive bidding. There is a symmetric Nash equilibrium in which every "working keyword" is allocated to a distinct firm type in X^*, and firms' bid-per-impression is equal to their average conversion rate, such that their surplus is fully extracted. On the other hand, the canonical broad-match function ensures that the consumers' search pools are diversified as mandated by the first-best. For this, it is crucial that when v is won by x-firms, $b(w|v)$ is proportional to $\sqrt{\mu(x,w)}$. The relation between the broad-match auction and the direct mechanism is quite transparent: when a firm submits the winning bid for $v \in W^*$, it effectively reports that its type is $f(v)$, and the allocation rule is determined by the broad-match function. As usual, the indirect mechanism generates more potential deviations, and therefore verifying equilibrium is more intricate.

The function p effectively assigns a *market equilibrium price* to every "working keyword". The Bhattacharyya coefficient plays an important role in the price function: as consumers' signals become less informative, the price-per-impression of all keywords in W^* weakly decreases. To illustrate the comparative statics, suppose $X = W$, the marginal of μ over X is uniform, and c is small such that $X^* = X$. Consider two extreme cases. First, suppose that $\mu(x|x) = 1$ for all x - i.e., consumers can perfectly describe their wants. Then, $p(x) = \theta$ for every x. Second, suppose that μ is uniform over $X \times X$. In this case, consumers' signals are entirely uninformative, and we have $p(x) = \theta / |X|$ for every x.

In general, the broad-match auction may give rise to symmetric pure-strategy Nash equilibria in which firms earn positive profits. The reason is the unusual non-competitive force that operates under this mechanism, noted in our discussion of the "classical music" example. Suppose that x-firms submit a winning bid for v only, whereas y-firms submit a winning bid for both v and v'. It is possible that y-firms will earn positive profits at v, and nevertheless refrain from raising their bid for v. The reason is that this deviation will crowd out x-firms from every search pool, and as a result consumers who search for x (and found a product they like in finite time prior to the deviation) will search indefinitely and bring the deviating firm's average conversion rate from v down to zero. The deviation will thus be unprofitable.

4.3 Discussion

From the consumers' point of view, the broad-match function plays the role of "vocabulary expansion". When a consumer submits the query w, he receives a search pool consisting of "tickets" that originate from various keywords. Therefore, the broad-

match function mimics an environment in which the consumer has a richer vocabulary. This feature was vividly demonstrated by our "classical music" example, where consumers who submit the generic query CL are linked (via the broad-match function) with the concrete keywords MOZ and STR, which in turn represent consumers who are able to give a precise description of their wants.

Real-life intermediaries regularly fulfill this vocabulary-expansion role, in order to generate matches that otherwise would not occur due to miscoordination in the language that parties on the two sides of the market use to describe themselves or what they are looking for. Our innovation is to show that when this role is *explicitly integrated into the design of the search platform*, it can augment a conventional (exact-match) competitive-bidding mechanism and go beyond its limitations.

Although our model borrows the term "broad matching" from the search-engine industry, it does not mean quite the same thing, and we believe that the differences are instructive.[6] Historically, broad matching was introduced in pursuit of two goals: (1) simplifying advertisers' bidding task, because under exact matching they need to bid for many queries (involving typos, synonyms or semantically related terms); (2) thickening markets that could be quite thin, thus raising the platform's auction revenues. Both goals are supply-side oriented - broad matching was not designed to fulfill the vocabulary-expansion role (although it effectively did). In practice, sponsored search coexists with "organic search"; the vocabulary-expansion role does not rely solely on broad matching in sponsored search, but on explicit attempts by the organic search engine to help the user refine and correct his query.

In our model, goals (1) and (2) are irrelevant: firms have unlimited ability to bid for as many keywords as they wish, and since there are infinitely many firms of each type, the markets would be thick even in the absence of broad matching. And of course, in our model the design of consumers' search environment is entirely decentralized, hence there is no "organic search engine" that can help consumers refine their queries. Thus, broad matching in our model is an exclusive and crucial instrument of vocabulary expansion.

As to the way broad matching interacts with the auction design, our "tickets" mechanism differs from current practice in two important qualitative dimensions. First, in our mechanism broad matching is not optional but forced upon bidders. In industry practice, broad matching is optional: bidders can choose to stick with exact matching.

[6] The following discussion is informed by discussions with David Pennock (a pioneer of sponsored-link auction formats, which employed broad matching) and Justin Rao. We are grateful to them for sharing their knowledge and insights, and apologize for any imprecision in our own description.

Second, in our model the winners in the auction for v are broad-matched with other queries w. In practice, broad matching means that whenever a firm bids for v and selects broad matching, the firm's bid for v enters as a bid in the auction for other queries w - thus contributing to the market-thickening objective. (Of course, there are other obvious differences: our model assumes unordered search; firms in our model know b whereas in practice the broad-match function is opaque; etc.)

Our motivation in this paper was emphatically not to provide an accurate model of how broad matching is implemented in the search-engine industry. Nevertheless, we believe that the perspective we provide - in particular, the idea that broad matching mimics the vocabulary-expansion role of search intermediaries (including "organic" online search engines) - may be relevant for future work on keyword auctions with broad matching (the current literature is small - see Section 6).

5 Ordered Search Pools

In this section we consider a search-design problem in which the designer can perfectly control the order by which consumers inspect search results. Thus, instead of having the consumer draw firms *at random* from a search pool, the designer now *chooses* which type of firm the consumer will encounter at each draw, as a function of his search history. For expositional simplicity, we assume in this section that c is sufficiently low, such that consumer surplus is maximized when all consumer types engage in search.

Consider an omniscient planner who can verify firms' product types. Such a planner would use its direct knowledge of firms' types to fix the exact order. The sequence of product types that maximizes the surplus of $(\cdot, w)$ consumers would be determined according to a simple maximum-likelihood calculation. The first type of product to be displayed, denoted $x_1(w)$, would most likely be the consumer's favorite conditional on his signal w - i.e., $x_1 \in \arg\max_{x \in X} \mu(x, w)$ (when some consumer types do not search, the likelihood is calculated for the set of participating consumer types). In general, the product type displayed in the k-th position of a w consumer's list, denoted $x_k(w)$, would be a product type most likely to be preferred by such a consumer, conditional on him not transacting with any of the $k-1$ firms whose types are $x_1(w), \ldots, x_{k-1}(w)$ (when θ tends to zero, the k-th product type on the list would simply be the k-th most popular product among w consumers). To simplify the exposition, assume that for each k there is a unique product type $x_k(w)$. An omniscient planner can also extract the entire surplus of firms by simply charging each firm a price per-impression that equals its conversion rate.

Consider now a search designer who *cannot* verify the firms' types. We argue that for *any* distribution of consumer types, the designer can implement the efficient sequence of firm types for every w, as well as fully extract firms' surplus, using the following anonymous direct mechanism. Each firm independently reports his type to the planner (it can also opt out from the mechanism and earn zero). As in Section 3, consumers do not participate in the mechanism but decide (on the basis of rational expectations) whether to search at the same time that firms report to the mechanism. For every profile of reports, the mechanism allocates each firm (probabilistically) to positions in each search pool, and specifies a (possibly negative) transfer that the firm pays to the search designer. As before, we take it for granted that all consumers search, as mandated by the first-best when c is small.

The mechanism's allocation rule is as follows. For each signal w, and for each position k in the list shown to a consumer who submits w, an individual firm is randomly drawn from those whose reported type is $x_k(w)$ (as defined above). The consumer will inspect a firm in the k-th position if and only if he does not transact with any of the first $k-1$ firms on his list. A firm that reports a type x pays a transfer equal to the expected number of transactions it will make, given the allocation rule and the assumption that all firms participate and report truthfully.

Let us now show that universal participation and truth-telling constitutes a Nash equilibrium. First, the transfer scheme means that all firms earn zero profits. Second, suppose that an individual x-firm considers reporting that its type is y. This will give the firm access to positions on various lists, which are meant to be allocated to y-firms. By construction, the consumer is more likely to want y rather than x, conditional on reaching each of those positions. Moreover, for each list that is displayed to a consumer, and for each position on that list, there is a measure 1 of firms from which the mechanism randomly draws one. It follows that an individual x-firm will obtain a lower expected number of transactions from the positions that are meant for y-firms. Therefore, no individual firm has any incentive to pretend to be a different type.

It follows that consumer-optimality and full extraction of firms' surplus are implementable in symmetric pure-strategy Nash equilibrium, under this mechanism. There is also an indirect mechanism that achieves these twin objectives. For every signal and for every position in the pool associated with that signal, the designer can run an independent first-price auction, where firms submit bids-per-impression, and the position is randomly allocated to one of the highest bidders.

The lesson from this section is perhaps that our search design problem is somewhat trivial when the planner can *fully* determine the order in which consumers inspect

alternatives. The problem appears to be more interesting when the planner has only an *imperfect* ability to control the order of inspection (for reasons that were listed in Section 2), or when the distribution of consumer preferences is richer.

6 Related Literature

We are not aware of precedent for our formulation of the "search design" problem. One related body of work is the literature on mechanisms for allocating sponsored links by online search engines. This literature (e.g. Edelman, Ostrovsky and Schwarz (2007)) mostly focuses on the mechanism-design problem of auctioning multiple "sponsored links". Typically, links are assumed to have exogenous values to advertisers. Athey and Ellison (2011) explicitly model how these values are determined by consumers' endogenous search decisions. Chen and He (2011) and Eliaz and Spiegler (2011) model explicitly the interaction between keyword and product prices (ignoring auction-theoretic considerations). Again, this literature almost invariably assumes exact matching.[7] Another important difference is that we assume a competitive environment with many firms of each type, whereas most of the literature on search engine pricing assumes small numbers of firms, such that auction-theoretic considerations become relevant.

Our discussion of competitive indirect mechanisms relates the paper to the literature on intermediation in two-sided markets (see Armstrong (2006), Caillaud and Jullien (2001,2003) and Rochet and Tirole (2003)). Some works within this tradition (e.g. Hagiu and Jullien (2011)) explicitly address search platforms. One can view the consumer's signal in our model as a (sole) platform to which he has access. In this sense, consumers are "single-homing" whereas firms are "multi-homing". Our key innovation in relation to this literature is the introduction of broad matching, which is essentially formalized as a "*directed network of platforms*". The papers we are aware of implicitly assume exact matching; multiple platforms are considered only in the context of competition among platforms, and interaction between a consumer and a firm invariably requires that they are both attached to the same platform.[8]

[7] Two exceptions are Dhangwatnotai (2011) and Chen et al. (2014). The first study uses the "price of anarchy" framework to analyze the performance of a mechanism in which advertisers can submit a bid to multiple generalized second-price auctions at once. The second paper analyzes the worst equilibrium of a mechanism that randomly samples a keyword according to a predefined probability distribution and only runs the generalized second-price auction for this sampled keyword.

[8] Some papers in the literature on two-sided markets have taken a mechanism-design approach. For instance, Spiegler (2000) examines contract design by an intermediary who aims to match agents who could interact elsewhere, and extract their joint surplus. Gomes and Pavan (2014) study mechanisms for implementing efficient many-to-many matching when agents are privately informed about their "vertical" characteristics.

Finally, in the last decade there has been much writing, both academic and popular, about the "long tail" phenomenon (see Brynjolfsson et al. (2006) or Anderson (2007)). This refers to the fact that tastes for many kinds of products are highly differentiated, such that a large segment of the consumer population belongs to a large number of small taste niches. The literature makes the observation that online commerce facilitates the flourishing of firms that serve the "long tail" because it lowers barriers that characterize brick-and-mortar commerce (such as storage costs). The key friction that remains (and possibly gets magnified) in such environments seems to be consumers' limited awareness of products that cater to their particular tastes, and limited ability to describe these tastes in order to locate relevant products on the internet. The "long tail" phenomenon means that the welfare implications of well-designed search environments can be large.

7 Conclusion

This paper addressed the following general question: *under what conditions can a decentralized mechanism be efficient in helping individuals find what they want?* In our leading example and in many of our discussions, we interpreted consumers' signals as keyword-based queries, and this linked our model to the problem of search-engine design. In this context, our question could be rephrased as follows: *suppose that a search intermediary who wishes to maximize the effectiveness of consumer search switches from a centralized matching algorithm to a decentralized mechanism (in particular, one based on competitive bidding for keywords); will search quality deteriorate as a result?*

However, our framework accommodates a wider range of environments, including ones which have yet to establish an organized marketplace for allocating firms to search pools. For example, online recommender systems give individuals access to search pools according to their past online activities, which serve as imperfect signals of their current needs. In contrast to search engines, recommender systems no not purely rely on queries initiated by the web user. For instance, Netflix automatically displays movie recommendations for its subscribers on its homepage; when a consumer buys a particular product on Amazon, the checkout screen displays recommended products, even though the consumer was not actively searching for these products; and when a researcher views a scholarly article on ScienceDirect, the side panel displays links to other recommended articles.

To see how our broad-match auction accommodates the recommendation-system interpretation, suppose that W represents a set of possible *past purchase profiles* of the consumer. In particular, we can set $W = X^K$, where K is the number of past purchase

opportunities the consumer had. A profile of past purchases serves as a platform for "personalized advertising", which is augmented by our notion of broad matching. Thus, when an advertiser pays for a particular profile of past purchases, he gets access to some set of profiles. In this context, our question can be rephrased as follows: *suppose that a recommender system abandons its centralized recommendation algorithm in favor of a "market for sponsored recommendations"; will the quality of its recommendations necessarily deteriorate as a result?* Our main result means that in the presence of small popularity gaps or strong correlation between past purchases and present tastes, a market-based recommendation system can theoretically mimic an ideal centralized recommendation algorithm. The same insight holds for the interpretation of w as a set of "cookies", namely passive indicators of the consumer's preferences (e.g. navigation history).[9]

The algorithms used by centralized recommender systems often rely on so-called *"topic models",* which are statistical tools employed by machine-learning specialists for inferring a latent abstract "topic" or "theme" that characterizes objects in a certain class (for a survey of these models see Blei and Lafferty (2009)).[10] For instance, suppose that an object is a scientific paper, the description of which is reduced to its frequency distribution of words. The idea is that different topics tend to generate different word distributions - e.g., a decision-theory paper will tend to have a greater frequency of the cluster of terms "utility", "Independence" and "Hausdorff". The machine-learning problem is to estimate a joint distribution μ over $X \times W$, where X is the set of papers (reduced to their word frequencies) and W is a set of latent topics whose size is fixed a priori, where $|W| \ll |X|$. After estimating the distribution, the recommender system often applies Bhattacharyya similarity to evaluate whether two papers have similar conditional topic distributions, and uses this judgment to make recommendations ("if you were interested in paper x, you might also be interested in paper y"). We find it interesting that the same measure of similarity arises in our models from entirely different considerations: minimizing consumers' search costs and satisfying firms' incentives.

[9] For a model of cookie pricing, see Bergemann and Bonatti (2013).

[10] We thank Stephen Hansen for pointing out the connection to topic models.

References

[1] Anderson, C. (2008): *The long tail: Why the future of business is selling less of more*, Hyperion Books.

[2] Armstrong, M. (2006): "Competition in two-sided markets," *Rand Journal of Economics* 37(3), 668-691.

[3] Athey, S. and G. Ellison (2011): "Position Auctions with Consumer Search," *Quarterly Journal of Economics* 126(3), 1213-1270.

[4] Basu, A. H. Shioya and C. Park (2011): *Statistical Inference: The Minimum Distance Approach*, CRC Press.

[5] Bergmann, D. and A. Bonatti (2013): "Selling Cookies," Cowles Foundation Discussion Paper No. 1920.

[6] Bhattacharyya, A. (1943): "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society* 35, 99–109.

[7] Blei, D. and J. Lafferty (2009): "Topics Models," in A. Srivastava and M. Sahami, editors, *Text Mining: Classification, Clustering, and Applications* . Chapman & Hall/CRC Data Mining and Knowledge Discovery Series

[8] Brynjolfsson, E., Yu J., and M. Smith (2006): "From Niches to Riches: Anatomy of the Long Tail." *Sloan Management Review* 47(4), 67-71.

[9] Caillaud, B. and B. Jullien (2001): "Competing Cybermediaries," *European Economic Review* 45(4), 797-808.

[10] Caillaud, B. and B. Jullien (2003): "Chicken and Egg: Competition Among Intermediation Service Providers," *Rand Journal of Economics* 34(2), 309-328.

[11] Chen, Y., C. He (2011): "Paid Placement: Advertising and Search on the Internet," *Economic Journa*l 121(556), F309-F328.

[12] Chen, W., D. He, T-Y. Liu, T. Qin, Y. Tao and L. Wang (2014): "Generalized Second Price Auction with Probabilistic Broad Match." *Proceedings of the 15th ACM conference on Electronic Commerce*, 39-56.

[13] Cormen, T. H., Stein , C., Rivest, R. L. and C. E. Leiserson (2001): *Introduction to Algorithms,* McGraw-Hill Higher Education.

[14] Dhangwatnotai, P. (2011): "Multi-keyword sponsored search," *Proceedings of the 12th ACM conference on Electronic Commerce*, 91-100.

[15] Edelman, B., M. Ostrovsky and M. Schwarz (2007): "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords." *American Economic Review* 97(1), 242-259.

[16] Eliaz, K. and R. Spiegler (2011): "A Simple Model of Search Engine Pricing," *Economic Journal* 121(556), 329-339.

[17] Gomes, R. and A. Pavan (2014), "Many-to-Many Matching and Price Discrimination", mimeo.

[18] Hagiu, A. and B. Jullien (2011): "Why Do Intermediaries Divert Search?," *Rand Journal of Economics* 42(2), 337-362.

[19] Rochet, J-C. (1987): "A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context," *Journal of Mathematical Economics* 16(2), 191-200.

[20] Rochet, J-C. and J. Tirole (2003): "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association* 1(4), 990-1029.

[21] Spiegler, R. (2000): "Extracting Interaction-Created Surplus," *Games and Economic Behavior* 30(1), 142-162.

[22] Theodoridis. S. and K. Koutroumbas (2008), *Pattern Recognition*, 4th Edition, Academic Press.

[23] Vohra, R. (2011): *Mechanism Design: A Linear Programming Approach*, Econometric Society Monograph, Cambridge University Press.

Appendix: Proofs

Remark 1

Denote $\mu(j|i) = \beta_{ik}$, such that (β_{ik}) is a stochastic matrix with $\Sigma_k \beta_{ik} = 1$ for every i. Let

$$\delta_{ik} = \sum_h \beta_{ih} m_{hk}$$

where (m_{hk}) is a $|W| \times |W|$ bi-stochastic matrix. Thus, (δ_{ik}) is a Blackwell garbling of (β_{ik}). Fix i, j. Then,

$$\sum_k \sqrt{\delta_{ik}\delta_{jk}} = \sum_k \sqrt{\left(\sum_h \beta_{ih} m_{hk}\right)\left(\sum_h \beta_{jh} m_{hk}\right)}$$

By the Cauchy-Schwarz inequality, this expression is weakly greater than

$$\sum_k \sum_h \sqrt{\beta_{ih} m_{hk} \beta_{jh} m_{hk}} = \sum_h \sqrt{\beta_{ih}\beta_{jh}} \sum_k m_{hk} = \sum_k \sqrt{\beta_{ik}\beta_{jk}}$$

Since this inequality holds for every i, j, it follows that

$$\sum_i \sum_k \sqrt{\delta_{ik}\delta_{jk}} \geq \sum_i \sum_k \sqrt{\beta_{ik}\beta_{jk}}$$

which completes the proof.

Lemma 1

Assume the contrary, namely that there exist w, x, y such that $\mu(y, w) < \mu(x, w)$ but $\lambda^*(y, w) > \lambda^*(x, w) = 0$. Consider switching from λ^* to λ', where the only difference is that $\lambda'(y, w) = 0$ and $\lambda'(x, w) = \lambda^*(y, w)$. This changes consumer surplus by the following amount

$$[\mu(x, w) - \mu(y, w)][1 - \frac{c}{\theta \cdot \lambda^*(y, w)}]$$

Since $\lambda^*(y, w) > c/\theta$, the change is positive, a contradiction.

Proposition 1

The surplus of consumers whose signal is w can be written as follows:

$$V(m) \equiv \sum_{i=1}^{m} \mu_i (1 - \frac{c}{\theta} \cdot \frac{\sum_{j=1}^{m} \sqrt{\mu_j}}{\sqrt{\mu_i}})$$

For any $m \in \{1, \ldots, |X|\}$,

$$V(m) - V(m-1) = \mu_m(1 - \frac{c}{\theta}) - \frac{2c}{\theta}\sqrt{\mu_m}\sum_{i=1}^{m-1}\sqrt{\mu_i} \tag{24}$$

This equation illustrates the negative search externality that consumer types exert on each other. The first term on the R.H.S represents the welfare gain when consumers who like product type m are added to the search pool. The second term represents the welfare loss due to the search costs incurred by the marginal consumer as well as the added search costs that he inflicts on other consumers (they now search longer since sometimes they draw m products).

Type m is the cutoff type if $V(i)-V(i-1) > 0$ for every $i \leq m$ and $V(i)-V(i-1) < 0$ for every $i > m$. Notice that as m increases, μ_m decreases while $\sum_{i=1}^{m-1}\sqrt{\mu_i}$ increases. Hence, if the R.H.S. is negative for some m, it is also negative for any $m' \geq m$. It follows that there exists a maximal index $m \in \{1, \ldots, |X|\}$ for which $V(m) - V(m-1) > 0$. By (24), this index, denoted m^*, satisfies that for any consumer type with $\mu_m \geq \mu_{m^*}$,

$$\sqrt{\mu_m} > \frac{2c}{\theta - c}\sum_{i=1}^{m-1}\sqrt{\mu_i}$$

while this inequality is reversed for any consumer type with $\mu_m < \mu_{m^*}$.

Proposition 5

Let us first verify that the strategy profile induces λ^*. Recall there is a measure one of every firm type $x \in X^*$. Since all x-firms bid exclusively for $f^{-1}(x)$, the measure of tickets held by x-firms in the search pool associated with any w is $b^*(w|f^{-1}(x))$. Therefore, for every $w \in W$,

$$\lambda(x, w) = \frac{b^*(w|f^{-1}(x))}{\sum_{y \in X^*} b^*(w|f^{-1}(x))}$$

Plugging the definition of b^*, we obtain

$$\lambda(x, w) = \frac{\sqrt{\mu^*(x, w)}}{\sum_{y \in X^*}\sqrt{\mu^*(y, w)}} = \lambda^*(x, w)$$

Our task now is the check that there are no profitable deviations. First, it is clear that no firm would want to bid for any $v \notin W^*$, because the assumption that $b(w|v) = 0$ for every such v and every $w \in W$ means that bidding for v would generate no transactions. Second, the strategy profile implies that if a consumer who submits

the query w draws a ticket held by an x-firm, then the ticket must be a $(w, f^{-1}(x))$ ticket. In other words, all the tickets held by x-firms in the search pool associated with w originate from the signal $v = f^{-1}(x)$, via the broad match link $b(w|v)$. This means that if a firm of type $y \neq x$ deviates by bidding $p > p(f^{-1}(x))$ for $f^{-1}(x)$, it would hold all the $(w, f^{-1}(x))$ tickets in some search pool w for which $\mu^*(x, w) > 0$. But this means that consumers of type (x, w) will search indefinitely without finding a product they like, and therefore the y-firm will get infinitely many impressions at the search pool associated with w. As a result, the average conversion rate that the y-firm will experience thanks to winning $f^{-1}(v)$ will drop to zero, and the firm's deviation will be unprofitable.

It follows that the only deviations we need to examine are of two types: (I) a firm of some type $y \in X^*$ submits $p(v)$ for some $v \in W^*$ for which $f(v) = x \neq y$; (II) a firm of some type $x \in X^*$ changes its bid for $v = f^{-1}(x)$ to $p > p(v)$.

Consider type-I deviations. Because the deviation is by a non-atomic firm that joins a measure-one set of highest bidders for v, there is a measure one of firms that hold all (w, v) tickets both before and after the deviation. Therefore, all we need to do is verify that the deviating y-firm generates fewer transactions than x-firms thanks to submitting the winning bid for v. The number of transactions for the deviating y-firm is

$$\sum_w \frac{b^*(w|v)}{1} \cdot \frac{K\mu^*(y, w)}{b^*(w|f^{-1}(y))} \tag{25}$$

because the firm gets $b(w|v)/1$ tickets to any pool w. In every such pool, there is a measure $K\mu^*(y, w)$ of consumers who are interested in y. Every such consumer eventually finds a product he likes, and this clientele is shared equally by all y-firms in the pool, the measure of which is $b^*(w|f^{-1}(y))$. By a similar calculation, the number of transactions for an x-firm is

$$\sum_w \frac{b^*(w|v)}{1} \cdot \frac{K\mu^*(x, w)}{b^*(w|v)} \tag{26}$$

Note that the last expression is equal to $K\mu^*(x)$. We need this quantity to be weakly greater than (25). Plugging the definition of b^*, we obtain

$$\mu^*(x) \geq \sum_w \sqrt{\mu^*(x, w)\mu^*(y, w)}$$

which is equivalent to the inequality (18). Thus, the condition that prevents type-I deviations coincides with the condition for general implementability of our twin

objectives, which was assumed to hold at the outset.

Before we check type II deviations, let us first check that all firms earn zero profits in equilibrium. Equivalently, we need to check that for every $v \in W^*$, $p(v)$ is equal to the average conversion rate experienced by an individual $f(v)$ firm thanks to submitting a winning bid for v. We saw that the number of transactions such a firm expects is $K\mu^*(f(v))$. The number of draws it obtains is

$$\sum_w b^*(w|v) \sum_{y|\mu^*(y,w)>0} \frac{K\mu^*(y,w)}{\theta \cdot \lambda^*(y,w)} \cdot \frac{1}{\sum_{v'} b^*(w|v')}$$

After substituting the expressions for b^* and λ^* and simplifying, this expression becomes

$$\frac{K}{\theta} \sum_y \sqrt{\mu^*(f(v))\mu^*(y)S^*(f(v),y)}$$

The ratio between the number of transactions and the number of draws, which is equal to $p(v)$. Hence, the average conversion rate is equal to the price per impression, which means that the firm earns zero profits.

let $f(v) = x$ for some $v \in W^*$, and suppose that an x-firm deviates by submitting a bid $p > p(v)$. The firm receives all the (w, v) tickets. However, both before and after the deviation, all (w, v) tickets are held by x-firms. Therefore, the composition of all the search pools remains intact, and therefore the number of transactions and the number of draws per every (w, v) ticket remains unchanged. It follows that the average conversion rate associated with v remains $p(v)$, which is below the price per impression that the deviating firm pays. Therefore, the firm's deviation is unprofitable.