Broad Matching and the Market for Search Platforms*

Kfir Eliaz[†] and Ran Spiegler[‡]

June 29, 2014

Abstract

We construct a two-sided market model, in which consumers can only provide a noisy signal of the type of product they want. Each signal functions as a platform, to which (multi-homing) firms get access if they pay its competitive market price. A "broad match" function, designed ex-ante by the search intermediary, links signals to one another: a firm that attaches itself to one signal (by paying its market price) can get access to the search pool of consumers who provide another signal. We ask the following question: *Is there a broad match function that induces an efficient market equilibrium, given the underlying search technology?* In the case of random sequential search, we provide a necessary and sufficient condition, in terms of the underlying joint distribution over consumers' tastes and signals - specifically, a simple inequality that involves the relative fractions of consumers who like different products, and the Bhattacharyya/Hellinger distance between their conditional signal distributions. The same inequality turns out to be the condition for joint implementability of efficiency and full surplus extraction under a general anonymous mechanism. The role that Bhattacharyya distance plays in our analysis links our paper to the machine-learning literature on recommender systems.

*This paper benefitted from an ISF grant, no. 1153/13. An earlier version was circulated under the title "The Market for Keywords". We thank Yair Antler, Yossi Azar, Alex Frug, Stephen Hansen, Sergiu Hart, seminar participants at Hebrew University, NYU, Technion, UPF, ESSET 2013 and the Pricing Workshop at Collegio Carlo Alberto, 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

A search platform is a site that attract firms from one side of a market into a "search pool" - a collection of firms with which consumers on the other side transact via some search process. A search intermediary (SI henceforth) is a market institution that provides such "search platforms". Real-life examples include human-resource or real-estate agencies, classified directories and, more modernly, online search engines and recommender systems.

The starting point of this paper is a simple theoretical observation: *a horizontally differentiated two-sided market in which firms compete for access to a single, undifferentiated search platform can fail to achieve an efficient outcome.* To see why, imagine that consumers can only survey a finite number of alternatives. Because firms' access to the platform is governed by market competition, only firms with the highest willingness to pay for access will get it. These are likely to be the firms that offer popular products; competitive forces will crowd out firms that cater to minority tastes. In other words, *the SI will fail to serve the "long tail" of the consumer preference distribution* (to use the terminology of Anderson (2007)).

Of course, SIs often gather information about consumers' preferences, in an attempt to "personalize" their search pools. For instance, when a prospective employer approaches an HR agency, he indicates the kind of worker he needs. In the case of classified directories, consumers consult the index in order to focus their search on a specific product category. Modern online platforms epitomize this tendency: search engines enable the consumer to submit an arbitrarily refined search query; and additional information about the consumer's preferences (past purchases, navigation history) is encapsulated in the "cookies" on his computer. Indeed, if the SI can obtain a perfect signal about the consumer's preferences, it can sort the two-sided market into homogeneous segments. In such a differentiated two-sided market, each signal functions as a distinct search platform, potentially with its own access price.

However, consumers rarely provide perfect signals about their true wants. A consumer's past purchase is clearly a noisy signal: if he ordered a vacation to Paris last summer, does this indicate that he likes big cities? Or will he want to diversify and try a beach resort this summer? In the case of a classified directory, consumers may struggle to fit what they look for into its rigid classification scheme. Even when consumers can submit any free-text query, they are effectively restricted by a limited ability to describe their wants. They may forget how to spell a name; they may be able to articulate only a general product category (e.g. movie genre); and their verbal descriptions may be vague, either due to inherent ambiguity (does "football" mean soccer or American football?) or because giving precise descriptions is hard ("the blonde singer who sounds like Rihanna"). Signal imperfection means that when the supply of each type of product is large, *the differentiated two-sided market will experience the same market failure we identified in undifferentiated ones* - namely, neglect of the "long tail" of the preference distribution associated with a given signal.

How do SIs cope with this predicament? A common practice is to take firms that attach themselves to a search platform associated with one signal, and introduce them into the search pool of consumers who are characterized by another signal. We refer to this device as *"broad matching"*, borrowing the terminology from online search engines. Indeed, when a web user submits a query to an online search engine, he gets a mixed collection of web links, which reflects the "semantic field" around the user's query, as well as an estimation of his underlying preferences. For instance, Googling "ninth symphony" produces a variety of links, referring to ninth symphonies by various composers, mostly Beethoven. Broad matching is common in "offline" settings as well. If a prospective buyer asks a real-estate agent for apartments in Downtown Manhattan, a sensible agent will provide properties listed by their owners only under "Nolita" (the name of a specific downtown neighborhood). Similarly, when a prospective employer asks an HR agent for a "junior sales manager", the agent may suggest a candidate who listed himself as an "experienced sales person".

We pose the following question: *Can a competitive, differentiated two-sided market implement an efficient outcome, under suitably designed broad matching?* And when it cannot, will another mechanism perform better? We construct a model of a two-sided market, in which firms complete for access to search platforms on one side, and consumers make search decisions on the other side. For most of the paper, we assume a random sequential search technology (without recall). For ease of exposition, we adhere to the concrete terminology of keyword search, such that a consumer's noisy signal represents his limited "vocabulary" for describing his wants, and the market for search platforms becomes a "*market for keywords*". Although this is suggestive of online search, our stylized model is not meant to be a faithful description of contemporary online search engine (and our "offline" examples imply that the notion of a "market for keywords" is not necessarily restricted to online search).[1] We discuss possible implications for other classes of search platforms in the concluding section.

[1]Our model departs from the current practice of online search engines in several dimensions. For instance, in reality online advertisers can *choose* between narrow and broad matching, whereas in our model the SI *imposes* the broad matching structure on market participants.

1.1 An Illustrative Example: Mozart vs. Stravinsky

Before presenting the model, we illustrate some of our main ideas with the following simple example. There are two *products*, Mozart and Stravinsky (denoted moz and str), each offered by a measure 1 of firms. Each consumer can provide a signal of his taste by means of a search query. Specifically, there are three *keywords*, "Mozart" (MOZ), "Stravinsky" (STR) and "Classical Music" (CL); and each consumer is characterized by a pair (x, w), where x is the (only) product he likes and w is the (only) word he can articulate in an attempt to describe what he is looking for. When a consumer transacts with a firm that offers his favorite product, both parties get a payoff of 1; otherwise, the transaction value is 0 for both parties.

The 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; while he can identify whether he likes a piece of music when he encounters it, all he can say ex-ante is that he is interested in classical music. Let $\mu(x, w)$ be the fraction of type (x, w) in the consumer population. Assume $\mu(moz, STR) = \mu(str, MOZ) = 0$ - that is, when a consumer can name a composer, then he must like his music. In addition, assume $\mu(moz, CL) > \mu(str, CL)$ - that is, moz is the popular composer among consumers whose vocabulary is CL. In this sense, moz is a "mass" product while str is a "niche" product.

When a consumer chooses to enter the market, he submits a query consisting of the single keyword he knows. The SI then provides him with a pool consisting of measures of moz and str firms. The search technology available to the SI is limited: it cannot impose an order of inspection on the elements in the consumer's search pool (in Section 2 we discuss justifications for this assumption). The consumer repeatedly draws random products from the pool without recall, and checks whether he likes them. Each draw - or "click", to use online-search terminology - carries a cost $c \in (0, 1)$. As soon as the consumer finds a product he likes, he transacts with the firm that offers it and terminates the search process; otherwise, he decides whether to continue searching.

How does the SI determine the composition of the consumer's search pool? An ideal centralized SI could directly identify the type of each firm. In contrast, Our SI *cannot* assign a keyword to firms according to some known relation between their products and the keyword's natural meaning. Instead, it subjects firms' access to market forces: *each keyword is offered at some price per draw* ("per click", to use search-engine terminology). Firms can pay for as many keywords as it wishes. Because consumer search is random without recall, firms compete for each draw independently, such that the market equilibrium price per draw will equal the highest willingness-to-pay (WTP) for it. The "meaning" of a keyword is determined *endogenously* in market equilibrium, according to the bundles of keywords that each firm chooses to pay for. We are interested in market equilibria that are "robust", in the sense that small perturbations of the distribution μ would not upset the market allocation of keywords to firms. This means that all firms of the same type must behave identically in equilibrium.

As is customary (indeed, taken for granted) in the literature on two-sided markets, suppose that the SI follows "narrow matching": a consumer who submits a query w receives only the firms that paid the market price of w. Assume that all consumers whose vocabulary is CL decide to search, and all firms of both types access CL as well. A firm's total WTP for CL is equal to the number of transactions it expects from consumers who submit the query CL. (To obtain the firm's WTP per draw, we simply need to divide total WTP by the expected number of draws. Since the latter is the same for all firms, regardless of their type, we can focus on the total WTP.) Since there is a measure 1 of *moz* and *str* firms, total WTP for CL will be $\mu(moz, CL)$ for a *moz* firm and $\mu(str, CL)$ for a *str* firm. And because $\mu(moz, CL) > \mu(str, CL)$, *moz* firms will win access to the draws of CL consumer, and crowd out *str* firms. It follows that we cannot sustain a robust equilibrium in which both (moz, CL) and (str, CL) are served.

When search costs are small, efficiency requires *all* consumer types to be served. Yet under narrow matching, the equilibrium with the highest social surplus allocates the keyword CL to *moz* firms, and (str, CL) consumers opt out (the keywords MOZ and STR are allocated to *moz* and *str* firms, respectively). The "market failure" of this equilibrium is thus that (str, CL) consumers - the "long tail" of the preference distribution associated with the keyword CL - do not get their desired product.

Can the SI use "broad matching" to overcome this market failure? Suppose that when a firm pays for STR (respectively, CL), it enters the search pool associated with the query CL with probability $b(CL|STR)$ (respectively, $b(CL|CL)$). Access to the search pools associated with MOZ and STR continues to be defined by "narrow matching": a firm gets access to the search pool associated with any of these two words if and only if it pays its market price. If *moz* firms are allocated to MOZ and CL, and *str* firms are allocated to STR (just as in the original equilibrium under narrow matching), a consumer who submits CL will get a search pool consisting of a measure $b(CL|CL)$ of *moz* firms and a measure $b(CL|STR)$ of *str* firms. Let $\lambda(x, w)$ denote the fraction of x firms in the search pool of consumers who submit the keyword w. Then,

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

This would allow both types (moz, CL) and (str, CL) to find what they like in finite time.

Recall that for small c, efficiency requires that all CL consumers engage in search. Moreover, an efficient composition of their search pool will minimize total search time:

$$\frac{\mu(moz, CL)}{\lambda(moz, CL)} + \frac{\mu(str, CL)}{\lambda(str, CL)}$$

By first-order conditions, the socially optimal composition is

$$\lambda(str, CL) = \frac{\sqrt{\mu(str, CL)}}{\sqrt{\mu(str, CL)} + \sqrt{\mu(moz, CL)}} \tag{2}$$

It follows that if we want the "broad matching" of CL consumers to induce an efficient outcome, we need to equate (2) and (1), thus obtaining the equation

$$\frac{b(CL|STR)}{b(CL|CL)} = \frac{\sqrt{\mu(str, CL)}}{\sqrt{\mu(moz, CL)}} \tag{3}$$

Is this equation consistent with (robust) market equilibrium? It turns out that *the introduction of broad matching creates a new incentive problem that did not exist under narrow matching*. To see why, note that the broad-match link from STR to CL means that if a *moz* firm pays for STR, it potentially gets access to consumers who submit CL, and *moz* is the popular product among this group. As a result, *moz firms may have a higher WTP for the keyword STR than str firms*. If this is the case, competitive forces will lead *moz* firms to crowd out *str* firms from STR!

This is the essence of the incentive constraint pertaining to the design of a broad match function that sustains an efficient equilibrium in the market for keywords. On one hand, broad matching addresses the "long tail" market failure resulting from consumers' limited ability to describe their wants, and increases the variety of products available to these consumers. Yet on the other hand, indiscriminate use of broad matching may encourage firms selling mass-appeal products to overtake too many keywords, thus exacerbating the "long tail" market failure.

To ensure that the original allocation of keywords to firms is consistent with market equilibrium under broad matching, we need to check that no single *moz* firm would be able to afford the keyword STR. Let us first calculate the WTP of an individual moz firm for STR. With probability $b(CL|STR)$, the firm would get access to a measure $\mu(moz, CL)$ of consumers who want the firm's product. However, the firm will have to share these consumers with a measure $b(CL|CL)$ of moz firms that paid for CL. It follows that the number of transactions that a single moz firm expects from STR is

$$\frac{b(CL|STR) \cdot \mu(moz, CL)}{b(CL|CL)} \tag{4}$$

Similarly, the number of transactions that a single str firm expects from STR is

$$\frac{b(STR|STR) \cdot \mu(str, STR)}{b(STR|STR)} + \frac{b(CL|STR) \cdot \mu(str, CL)}{b(CL|STR)} = \mu(str) \tag{5}$$

where $\mu(x) = \sum_w \mu(x, w)$. It follows that moz firms will not pay the market price of STR in equilibrium if

$$\mu(str) > \frac{b(CL|STR) \cdot \mu(moz, CL)}{b(CL|CL)} \tag{6}$$

Inserting (3), condition (6) can be rewritten as

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

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

$$\frac{\mu(moz)}{\mu(str)} \cdot \left(\sum_w \sqrt{\mu(w|moz)\mu(w|str)} \right)^2 < 1 \tag{7}$$

The second multiplicative term on the L.H.S of (7) is a conventional measure of similarity between the conditional distributions $\mu(\cdot|moz)$ and $\mu(\cdot|str)$, known as "*Bhattacharyya similarity*" (after Bhattacharyya (1943)). Thus, sustainability of the efficient outcome in market equilibrium may be obstructed by a large popularity gap between moz and str, or by similar query distributions that characterize moz and str lovers (which is the case if the vocabulary of many consumers is CL).

Although we arrived at the final inequality by assuming a special form of broad matching (modifying narrow matching by adding the terms $b(CL|CL)$ and $b(CL|STR)$) and a particular allocation of keywords to firm types, *this is in fact the general necessary and sufficient condition for market implementability of an efficient outcome*, for any

X that contains moz and str, any W consisting of at least two keywords, and any μ for which efficiency requires $\lambda(moz, w)\lambda(str, w) > 0$ for some w. This will be our main result in this paper.

1.2 Related Literature

Our paper is related to the literature on intermediation in two-sided markets (see Armstrong (2006), Caillaud and Jullien (2001,2003), Rochet and Tirole (2003) and Spiegler (2000)). Some works within this tradition (e.g. Hagiu and Jullien (2011)) explicitly address search platforms. Like much of this literature, we assume single-homing on the consumers' side and multi-homing on the firms' side. Our key innovation in relation to this literature is the introduction of broad matching, which is essentially formalized as a "*directed network of platforms*". All the papers we are aware of implicitly assume narrow 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. The platform-network aspect of our model also relates it to the literature on buyer-seller networks. In this literature (see Kranton and Minehart (2001)), agents can only trade with linked partners. Typically studied questions are which networks are efficient and which networks emerge from agents' strategic link-formation decisions.

Another related strand involved models of keyword pricing. This literature (e.g. Edelman, Ostrovsky and Schwarz (2007)) mostly focuses on the mechanism-design problem of auctioning multiple "sponsored links". Typically, the 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 (2011a) model explicitly the interaction between keyword and product prices (ignoring auction-theoretic considerations). Again, this literature almost invariably assumes narrow matching (see Dhangwatnotai (2011) for an exception). 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.

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)), namely 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, and 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 such tastes in order to locate relevant products on the internet. The "long tail" phenomenon means that the welfare implications of well-designed broad matching can be large.

2 The Model

Products and words

Let X be a finite set of *product types*. Denote $|X| \geq 2$. Let W be a finite set of *words*, where $|W| \geq |X|$ (we use the terms "word" and "keyword" interchangeably). 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 (only) word he can use to express his wants. We refer to w as the type'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 $\mu(\cdot|x) = (\mu(w|x))_{w \in W}$. We sometimes refer to the latter as the conditional query distribution that characterizes the preference type x.

A consumer of type (x, w) gets a payoff of 1 (0) with independent probability q $(1 - q)$ if he consumes a product of type x, and a sure payoff of 0 if he consumes a product of type $y \neq x$. Products are "inspection goods": when a consumer sees a product, he immediately recognizes the payoff it generates. For every $x \in X$, there is a measure one of *firms* that offer only that product type (as many units as required). A firm gets a payoff of 1 from any unit it sells (we abstract from product prices).

The parameter q captures idiosyncratic heterogeneity among consumers and firms. Each consumer is interested in one *type* of product, but for each product of this type there is an independent probability q that he will like it. This additional dimension of differentiated taste also justifies why many firms offer the same type of product.

Let $b : W \times W \to [0, 1]$ be a weighted directed graph over words, referred to as the "*broad match function*". This object is designed by the SI ex-ante. We use the following notation: $b(w|v)$ is the weight of the link from v to w (to avoid misunderstandings, we do *not* require $\Sigma_w b(w|v) = 1$). When $b(w|w) = 1$ and $b(w|v) = 0$ for all $w \neq v$, we refer to b as the *narrow match* function. When $b(w|v) = 1$ for all w, v, we refer to b as the *fully broad match* function.

Market equilibrium

Let $f : W \to X$ be an allocation of words to product/firm types, and denote $n_f(x) = |f^{-1}(x)|$. Let $a : X \times W \to \{0,1\}$ be a function that indicates the decision of each consumer type whether to engage in active search. Every pair f, a is endowed with two functions:

(i) $\pi_{f,a}(x, w)$ is the *number of transactions* that an individual x firm expects when it gets access to consumers whose vocabulary is w;

(ii) $t_{f,a}(x, w)$ is the *expected search time* for consumer type (x, w) if he submits the query w.

Consumers' search cost is $c \in [0, q)$ per time unit. For now, we take the functions $\pi_{f,a}$ and $t_{f,a}$ as primitives - later on we will give them a structure that reflects consumers' search process.

Definition 1 *The pair (f, a) is a* ***market equilibrium*** *if the following conditions hold:*
(i) For every $(x, w) \in Supp(\mu)$, $a(x, w) = 1$ if and only if $c \cdot t_{f,a}(x, w) < 1$.
(ii) If $f(v) = x$, then

$$\sum_w b(w|v)\pi_{f,a}(x, w) > \sum_w b(w|v)\pi_{f,a}(y, w)$$

for every $y \neq x$.

Condition (i) captures the individual rationality of search decisions: each consumer engages in search if and only if the expected cost search is below the gross payoff from finding a product he likes. Condition (ii) is a "market clearing" property: each word w is allocated to the firm type that values it the most. The reason we impose a strict inequality is that we want the equilibrium allocation to be stable w.r.t small perturbations of μ.

Given a market equilibrium (f, a), we define the *equilibrium access price* of the word v to be

$$p^*_{f,a}(v) = \sum_w b(w|v)\pi_{f,a}(f(v), w) \tag{8}$$

Thus, the equilibrium access price of a word is equal to the highest willingness to pay for it, in line with the idea of a competitive market for keywords. Later on, we will be interested in the equilibrium price-per-draw (or "price-per-click", to use the language of online search engines) of each word.

The search process: Defining $\pi_{f,a}$ and $t_{f,a}$

Definition 1 is stated for arbitrary functions $\pi_{f,a}, t_{f,a}$. We will now define them in terms of the primitives μ, b, in way that reflects a conventional *random sequential search technology.* When a consumer submits a query, he gets access to a *search pool.* He repeatedly draws independent random samples from this pool, and his search is terminated as soon as he finds a product he likes. Let us now see how to derive expressions for $\pi_{f,a}, t_{f,a}$ from such a search process.

First, note that given a, the total measure of consumers who demand the product x and submit the query w is

$$d_a(x,w) \equiv \mu(x,w)a(x,w)$$

Given f and b, the total measure of x firms that are available to consumers who submit w is

$$m_f(x,w) \equiv \sum_{v \in f^{-1}(x)} b(w|v) \tag{9}$$

It follows that the fraction of x firms in the search pool associated with the query w is

$$\lambda_f(x,w) \equiv \frac{m_f(x,w)}{\sum_{y \in X} m_f(y,w)} \tag{10}$$

The stopping probability per draw of consumer type (x,w) is $q \cdot \lambda_f(x,w)$. His expected search time (where a unit of time is one draw) is the inverse of this expression, hence

$$t_{f,a}(x,w) = \frac{\sum_{y \in X} \sum_{v \in f^{-1}(y)} b(w|v)}{q \cdot \sum_{v \in f^{-1}(x)} b(w|v)} \tag{11}$$

Because the consumer's payoff is 1 if he gets a product he likes and 0 otherwise, the consumer's optimal search decision is simple: either he he searches until he finds a product he likes, or he refrains from searching altogether. It follows every consumer who likes x and submits w eventually transacts with some x firm in the search pool associated with w. These consumers are equally shared by all x firms in the pool. Hence, whenever $m_f(x,w) > 0$, the number of transactions that an individual x firm obtains in the pool is $d_a(x,w)/m_f(x,w)$. When $m_f(x,w) = d_a(x,w) = 0$, we write $\pi_{f,a}(x,w) = 0$. Condition (i) in Definition 1 rules out the possibility that $d_a(x,w) > m_f(x,w) = 0$. It follows that

$$\pi_{f,a}(x,w) = \frac{\mu(x,w)a(x,w)}{\sum_{v \in f^{-1}(x)} b(w|v)} \tag{12}$$

The functions $t_{f,a}$ and $\pi_{f,a}$ are thus *defined* by (11) and (12). If we plug these expressions into Definition 1, we have a complete definition of market equilibrium in terms of the exogenous elements X, W, μ, b, c, q.

Comment: The search technology

Our assumption of a random sequential search technology means that *the SI is unable to impose an order of inspection on consumers' search pools*. This seems to fit well environments in which inspection is done "offline". Consider our HR-agency example; even if the HR agent is able to provide an ordered list of candidates for the prospective employer, the eventual order of interviews is likely to be subjected to physical constraints beyond the SI's control. Even in the case of online search, web users may disobey the order in which links appear on their computer screen, for a variety of reasons: advertisers may use "obfuscation" tactics to attract the user's attention away from the suggested order; some links may be slow or broken; and the user may distrust the search engine's suggested order (see Athey and Ellison (2011) for a related discussion). From this point of view, our random-search assumption can be viewed as an extreme, *worst-case* analysis for the SI (which is also computationally cheaper than complete ordering of *all* the firms in the consumer's search pool according to the firms' market behavior). In Section 5 we examine the diametrically opposed case, in which the SI 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.

The Bhattacharyya coefficient

For any pair of products $x, y \in X$, define:

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

This is a measure of similarity between the two conditional query distributions $\mu(\cdot|x)$ and $\mu(\cdot|y)$. Technically, $\sqrt{S(x,y)}$ 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. In the statistics literature, $\sqrt{S(x,y)}$ is known as the *Bhattacharyya coefficient* that characterizes the distributions $\mu(\cdot|x)$ and $\mu(\cdot|y)$. A related concept is the *Hellinger distance* between distributions, given by $H^2(x,y) = 1 - \sqrt{S(x,y)}$ (see Basu, Shioya and Park (2011) and Theodoris and Koutroumbas (2008)). In the concluding section we discuss applications of this concept in machine-learning models of recommender systems, and how these are related to our model.

The stochastic matrix $(\mu(\cdot|x))_{x\in X}$ is a signal function in Blackwell's sense. This leads to the following observation.

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

Proof. 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. ■

Thus, as consumers' queries provide weaker signals of their preferences, the measure $S(x,y)$ weakly increases for all x,y.

Welfare

In order to explore the welfare limitations of the market for search platforms, we assume that the SI is benevolent and aims to maximize social welfare (we discuss profit maximization in Section 6). The domain of our social welfare function is $(\Delta(X))^W$, namely the set of all possible collections of search pools to which consumers have access. A search pool is characterized by its composition of firm types. Thus, an element in the domain $(\Delta(X))^W$ is $\lambda=(\lambda(x,w))_{x\in X,w\in W}$, where $\lambda(x,w)$ is the fraction of x firms in the search pool associated with the query w. In our market model, λ is induced by

f, a via the formulas (9) and (10). However, the origin of λ is of course irrelevant for the definition of social welfare.

Recall that in market equilibrium, the access price of a word is equal to the highest willingness to pay for it; as a result, firms earn zero profits in equilibrium. Therefore, we equate social welfare with consumer surplus, and define the social welfare function U as follows:

$$U(\lambda) \equiv \sum_{w} \sum_{x} \mu(x, w) u_\lambda(x, w), \tag{13}$$

where

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

is the net utility of consumer type (x, w) under (f, a). A market equilibrium (f, a) is *efficient* if $\lambda_{f,a}$, as defined by (10), maximizes social welfare.

Keyword prices

Condition (ii) in Definition 1 captures in a reduced form a (robust) competitive market allocation of keywords to firms. It is essentially a zero-profit condition. A more conventional definition would be based on an explicit description of the supply of "search space" provided by the SI and the firms' demand for it, and it would include an explicit market price for each word. The random-sequential-search search technology enables such an account. Suppose (as in Section 1.1) that the search process is without recall. Thus, every draw functions as an independent unit supply of "search space", which is competed for by a large number of firms of each type. The equilibrium *price-per-draw* clears the market when it is equal to the highest WTP per draw among firms. Because all firms face the same number of draws in each pool (conditional on getting access it), this is equivalent to our Condition (ii).

The following is an alternative scenario, which does not require the assumption of no recall. Imagine that access to search platforms is subjected to a physical capacity constraint: the maximal measure of firms that can be admitted to any w is less than 1. The market price of w (*a lump-sum payment for access, as opposed to a price-per-draw*) is $p(w)$. On the demand side, each x firm demands a bundle of words $A \subseteq W$ to maximize its total profit

$$\sum_{v \in A} \left[\sum_{w \in W} b(w|v) \pi(x, w) - p(v) \right]$$

where $\pi(x, w)$ is the number of transactions it expects in the search pool associated with w, given market agents' equilibrium behavior. In competitive equilibrium, $p(w)$ equates the supply and demand at every w (given the consumers' individually rational search decisions). If we require the equilibrium to be robust to small perturbations in μ, each word must be allocated to exactly one type of firms. As a result, this elaborate definition of market equilibrium boils down to Definition 1. It also justifies our definition of equilibrium access price given by (8).

Comment: Multi-product firms. The assumption that each firm sells one type of product is not essential. We could allow firms to sell multiple product types, as long as there are no search externalities between them. That is, a consumer's encounter with a product does not affect the probability of encountering another product offered by the same firm. In reality, firms sometimes use products as "baits" to lure the consumer into browsing through their entire product line (see Eliaz and Spiegler (2011b) for a stylized model of competitive marketing that captures this aspect).

Comment: Rational expectations. Condition (i) in Definition 1 assumes that consumers have rational expectations. This may seem strange: if the consumer is unable to express what he wants, how can he figure out the duration of searching for it? However, note that consumers often use the same generic word for many queries. Imagine that you heard a nice song on the radio, and the only thing you can say about it is that it is an R&B song. Therefore, this is the only keyword you can use to look for it on YouTube. Previous cases in which you submitted the query "R&B" on YouTube have enabled you to form an estimate of the expected search time. It is possible to relax the rational-expectations assumption, and simply assume that a consumer of type (x, w) submits the query w automatically, without performing any cost-benefit analysis. From a positive point of view, this would be equivalent to the case of $c = 0$ in our model. For the normative analysis, however, this alternative assumption raises the question of whether minimizing expected search time is a "legitimate" social welfare criterion, from a revealed-preference point of view.

3 Analysis

In this section we analyze market equilibria in our model, with particular emphasis on whether socially optimal outcomes can be sustained in market equilibrium for some broad match function. Our main results are in Section 3.2, where we characterize efficient outcomes under $c > 0$, and provide a necessary and sufficient condition for their market implementability.

3.1 The Limitations of Narrow and Fully Broad Match

Let us first examine the welfare implications of market equilibrium under the two extreme broad match functions: narrow and fully broad match.

Proposition 1 (optimal equilibrium under narrow match) *Let $b(w|w) = 1$ and $b(w|v) = 0$ for all $w \neq v$. Then, for generic μ, the maximal social welfare that can be sustained in market equilibrium is*

$$(1 - \frac{c}{q}) \sum_{w \in W} \left(\max_{x \in X} \mu(x, w) \right)$$

Proof. First, we construct a market equilibrium (f, a) that implements this level of social welfare. Let $f(w) = \arg\max_{x \in X} \mu(x, w)$. For generic μ, this is a well-defined function. Let $a(x, w) = 1$ if and only if $x = f(w)$. It is easy to see that both conditions of the definition of market equilibrium are satisfied. Moreover, the expected search time for consumers who choose to enter is $\frac{1}{q}$, which is the shortest possible. Hence their payoff is $1 - \frac{c}{q}$.

Now suppose there is another equilibrium (f, a). For each word w, let $X(w)$ be the set of products for which $a(x, w) = 1$. Then, the firm type with the highest WTP for w is $\arg\max_{x \in X(w)} \mu(x, w)$. For generic μ, this is a singleton, hence also $f(w)$. But this means that condition (i) in Definition 1 is satisfied only if $X(w)$ consists of a single element $x(w)$, such that social welfare is

$$(1 - \frac{c}{q}) \sum_{w \in W} \mu(x(w), w) \leq (1 - \frac{c}{q}) \sum_{w \in W} \left(\max_{x \in X} \mu(x, w) \right)$$

which completes the proof. ■

Thus, under narrow match, it is impossible to do better than serving the largest preference niche among consumers who share a given vocabulary. When $X \subseteq W$ and $\mu(x|x) = 1$ for all x - i.e. when consumers always know the name of the product they want - narrow matching enables the efficient equilibrium outcome. Narrow matching is also optimal when c is sufficiently close to q, such that serving the largest preference niche at each keyword is efficient.

Let us now turn to the diametrically opposed case of fully broad matching.

Proposition 2 (optimal equilibrium under fully broad match) *Let* $b(w|v) = 1$ *for all* w, v*. Then, for generic* μ*, the maximal social welfare that can be sustained in market equilibrium is* $(1 - \frac{c}{q}) \max_{x \in X} \mu(x)$*.*

Proof. Under fully broad match, the model becomes equivalent to a specification (W', μ', b'), where W consists of a single word w, $\mu'(x, w) = \mu(x)$, and b' is a narrow match function. By Proposition 1, the maximal social welfare that can be sustained in market equilibrium in this case is $(1 - \frac{c}{q}) \max_{x \in X} \mu(x)$. ■

The maximal social welfare that is implementable in market equilibrium is weakly lower under fully broad match than under narrow match. Narrow matching neglects the "long tail" of the preference distribution conditional on the consumer's vocabulary. However, going all the way to fully broad matching throws the baby with the bathwater, because it leads to a wholesale neglect of the long tail of the *unconditional* preference distribution.

3.2 Efficient equilibrium under $c > 0$

When $c > 0$, the social welfare function can be written as

$$U(\lambda) = \sum_{w} \sum_{x|\lambda(x,w)>0} \mu(x, w)[1 - \frac{c}{q \cdot \lambda(x, w)}] \tag{14}$$

Our first task is to characterize the collection of search pools $\lambda^* = (\lambda^*(x, w))_{x \in X, w \in W}$ that maximizes U. We perform this task in four steps.

First, it is immediately clear from (14) that we can calculate $(\lambda^*(x, w))_{x \in X}$ *independently for each* w.

Second, observe that if $\lambda^*(x, w) > 0$, then $\lambda^*(x, w) > \frac{c}{q}$. The reason is as follows. Recall that by assumption, $c < q$. Imagine that $\lambda^*(x, w) \leq \frac{c}{q}$, and suppose that we deviated by removing all x firms and all consumers of type (x, w) from the pool associated with w. This would weakly increase the payoff earned by (x, w) consumers. In addition, it would eliminate the negative search externality that x firms in the pool exert on consumers who like other products. It follows that *if* $\lambda_{f,a} = \lambda^*$*, then* (f, a) *must satisfy condition* (i) *in Definition 1.*

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

$$\frac{\lambda^*(x, w)}{\lambda^*(y, w)} = \sqrt{\frac{\mu(x, w)}{\mu(y, w)}} \tag{15}$$

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

$$\lambda^*(x,w) = \frac{\sqrt{\mu(x,w)}}{\sum_{y|\lambda^*(y,w)>0} \sqrt{\mu(y,w)}} \tag{16}$$

The fourth and last step characterizes the set of products 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 $\lambda^*(y,w) = 0$.*

Proof. 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 social welfare by the following amount

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

Since $\lambda^*(y,w) > \frac{c}{q}$, the change is positive, a contradiction. ■

This lemma has the following implication. For each word w, order the products 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 $\lambda^*(i,w) = \lambda_i^*$.) The efficient λ^* has the property that 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 efficient cutoff type is characterized as follows.

Lemma 2 *The cutoff m^* is the highest $m \in \{1, ..., |X|\}$ for which*

$$\sqrt{\mu_m} > \frac{2c}{q-c} \sum_{i=1}^{m-1} \sqrt{\mu_i} \tag{17}$$

Proof. Total consumer surplus among consumers whose vocabulary is w can be written as follows:

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

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

$$V(m) - V(m-1) = \mu_m (1 - \frac{c}{q}) - \frac{2c}{q} \sqrt{\mu_m} \sum_{i=1}^{m-1} \sqrt{\mu_i} \tag{18}$$

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 (18), this index, denoted m^*, satisfies that for any consumer type with $\mu_m \geq \mu_{m^*}$,

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

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

Equation (18) illustrates the negative search externality that consumer types exert on each other. The first term on the R.H.S represents the welfare gain for consumers who like product type m, when this type is 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).

Having characterized the efficient collection of search pools, our task now is to examine its implementability in market equilibrium. Recall that a market equilibrium (f, a) is efficient if $\lambda_f = \lambda^*$. Our next result provides a necessary and sufficient condition for the existence of a broad match function that induces an efficient equilibrium.

Proposition 3 (necessary and sufficient condition for efficient equilibrium) *There exists a broad match function b^* that induces an efficient market equilibrium (f, a) if and only if*

$$\frac{\mu(x)}{\mu(y)} S(x, y) < 1 \tag{19}$$

for every pair of distinct products x, y for which $\lambda^*(x, w)\lambda^*(y, w) > 0$ for some $w \in W$. In particular, f can be any function whose image is $\cup_{w \in W}\{x \in X \mid \lambda^*(x, w) > 0\}$, and b^* can be defined by

$$b^*(w|v) = \begin{cases} \frac{\sqrt{\mu(f(v),w)}}{n_f(f(v))} & if \quad \lambda^*(f(v), w) > 0 \\ 0 & if \quad \lambda^*(f(v), w) = 0 \end{cases} \tag{20}$$

Proof. Assume there is a broad match function b that induces an efficient equilibrium (f, a). Recall that if an individual x firm pays for v, it receives the following number of transactions:

$$\sum_w \frac{b(w|v)\mu(x, w)a(x, w)}{\sum_{v' \in f^{-1}(x)} b(w|v')}$$

where words w for which $\sum_{v' \in f^{-1}(x)} b(w|v') = 0$ are ignored in the summation. When $\sum_{v' \in f^{-1}(x)} b(w|v') > 0$, we have $\lambda_f(x,w) > 0$, and we have already noted that in this case, $a(x,w) = 1$. It follows that we can assume w.l.o.g that $a(x,w) = 1$ for every w in the summation. Condition (ii) in Definition 1 can thus be written as follows. For every $v \in W$ and every $y \neq f(v) = x$,

$$\sum_w \frac{b(w|v)\mu(x,w)}{\sum_{v' \in f^{-1}(x)} b(w|v')} > \sum_w \frac{b(w|v)\mu(y,w)}{\sum_{v' \in f^{-1}(y)} b(w|v')} \tag{21}$$

Summing Inequality (21) over all $v \in f^{-1}(x)$, we obtain

$$\sum_{v \in f^{-1}(x)} \sum_w \frac{b(w|v)\mu(x,w)}{\sum_{v' \in f^{-1}(x)} b(w|v')} > \sum_{v \in f^{-1}(x)} \sum_w \frac{b(w|v)\mu(y,w)}{\sum_{v' \in f^{-1}(y)} b(w|v')}$$

This simplifies into

$$\frac{\mu(x)}{\mu(y)} > \sum_w \left(\frac{\sum_{v \in f^{-1}(x)} b(w|v)}{\sum_{v \in f^{-1}(y)} b(w|v)} \right) \mu(w|y) \tag{22}$$

whenever $f(v) = x$. Similarly, we obtain

$$\frac{\mu(y)}{\mu(x)} > \sum_w \left(\frac{\sum_{v \in f^{-1}(y)} b(w|v)}{\sum_{v \in f^{-1}(x)} b(w|v)} \right) \mu(w|x) \tag{23}$$

whenever $f(v) = y$. Now, we can plug the definition of $\lambda_f(x,w)$ given by (9)-(10) and the necessary condition for efficiency given by (16) in inequalities (22) and (23), and obtain that the following inequalities must hold for every pair of distinct products x, y in the image of f:

$$\begin{aligned} \frac{\mu(x)}{\mu(y)} &> \sqrt{\frac{\mu(x)}{\mu(y)}} \sum_{w \in W} \sqrt{\mu(w|x)\mu(w|y)} \\ \frac{\mu(y)}{\mu(x)} &> \sqrt{\frac{\mu(y)}{\mu(x)}} \sum_{w \in W} \sqrt{\mu(w|x)\mu(w|y)} \end{aligned}$$

By the definition of $S(x,y)$ these inequalities may be rewritten as

$$\max\left(\frac{\mu(y)}{\mu(x)}, \frac{\mu(x)}{\mu(y)}\right) \cdot S(x,y) < 1$$

which implies the desired condition (19).

For the sufficiency part of the proof, note that by construction, the pair f, b^* yields the optimal λ^*. We have already noted that this immediately implies Condition (i) in Definition 1. Plugging the definition of b^* into (21) establishes Condition (ii). ■

Condition (19) captures in a succinct way the key considerations highlighted in Section 1.1. A high $S(x, y)$ captures an environment in which consumers' queries are weak indicators of their true wants. Broad matching is meant to address this problem, by giving consumers a diversified search pool. However, since words are allocated to firms via market competition, broad matching may increase the risk that a mass-appeal product will crowd out a niche product. This risk goes up as the popularity gap between the two products, captured by $\mu(x)/\mu(y)$, gets farther away from one.

Comment: Canonical b, f

Suppose that we impose the natural restriction $X \subseteq W$ - that is, the name of each product is itself a keyword. Then, it is also natural to impose two additional restrictions: $f(x) = x$ and $b(x|x) = 1$. The first restriction is w.l.o.g. Recall that by Proposition 3, we can select f to be *any* function whose image is $\cup_{w \in W}\{x \in X \mid \lambda^*(x, w) > 0\}$. If $\lambda^*(x, w) > 0$ for some w, setting $f(x) = x$ is consistent with this qualification. If $\lambda^*(x, w) = 0$ for all w, we can set $f(x) = x$ and $b(w|x) = 0$ for every w, and design f, b as if x were excluded from both X and W. The second restriction, however, carries a loss of generality: the condition for efficiency implies that setting $b(v|v) = 1$ may force the value of $b(w|v)$ to be greater than one for some w, a contradiction.

3.2.1 Equilibrium Keyword Prices

We now turn to a characterization of the access price of keywords in market equilibria. For simplicity, we make two restrictions. First, we assume that the broad match function is *symmetric*, in the sense that it does not discriminate between keywords that are allocated to the same product - i.e., $f(v) = f(v')$ implies $b(w|v) = b(w|v')$. The function b^* satisfies this property. Second, we focus on equilibria with full consumer participation: $a(x, w) = 1$ whenever $\mu(x, w) > 0$. This property holds in efficient equilibrium as long as c is sufficiently small.

Note that the L.H.S of condition (ii) of Definition 1 coincides with the definition of $p^*_{f,a}(v)$ given by (8). It follows that under our simplifying restrictions,

$$p^*_{f,a}(v) = \frac{\mu(f(v))}{n_f(f(v))}$$

This characterization of keyword prices does not rely on the efficiency of market equilibrium; it is a simple consequence of the feature that consumers in our model search until they find a product they like.

The characterization of equilibrium keyword prices turns out to be more interesting when we consider the average *price-per-draw* - or "per click", to use the terminology of online search. In some settings (e.g. online search), the SI can record consumers' visits at each firm, and thus charge firms per visit. (This method of payment may be preferable to lump-sum pricing when consumer traffic is uncertain - however, our model abstracts from this consideration.) Our reduced-form definition of market equilibrium does not involve an explicit notion of a market price of keywords, and it accommodates both methods of payment as consistent interpretations. When keyword prices are lump-sum payments for access, the notion of a price-per-draw is a fictitious "accounting" number.

The *number of draws* that any firm obtains in the search pool associated with w is

$$L_{f,a}(w) \equiv \frac{\sum_{x \in X} \frac{d_a(x,w)}{q \cdot \lambda_f(x,w)}}{\sum_{y \in X} m_f(y,w)} = \frac{1}{q} \sum_{x \in X} \pi_{f,a}(x,w)$$

The reasoning behind this expression is as follows. Since the stopping probability per draw of a consumer of type (x, w) is $q \cdot \lambda_f(x, w)$, he contributes $1/q\lambda_f(x, w)$ draws in expectation. The total number of draws by consumers is thus equal to the sum of $d_a(x, w)/q\lambda_f(x, w)$ over all products $x \in X$. These draws are uniformly distributed over all the firms in the pool, hence each firm gets a fraction $1/\sum_{y \in X} m_f(y, w)$ of the total number of draws.

Define the *conversion rate* of firm type x from the word v, induced by (f, a), as follows:

$$\begin{aligned} CR_{f,a}(x,v) &= \frac{\sum_w b(w|v)\pi_{f,a}(x,w)}{\sum_w b(w|v)L_{f,a}(w)} \\ &= \frac{q \cdot \sum_w b(w|v)\pi_{f,a}(x,w)}{\sum_{y \in X} \sum_w b(w|v)\pi_{f,a}(y,w)} \end{aligned}$$

To understand this expression, note that when an individual x firm is among the firms that were allocated the word v, it potentially enters multiple pools w, mediated by the broad match function. For each such pool, we can calculate the number of draws and transactions the firm can expect. The conversion rate is the ratio between the total (aggregated over all the search pools) numbers of transactions and draws.

Given a market equilibrium (f, a), the *price-per-draw* of the word v is defined as

$$PPD^*(v) = CR_{f,a}(f(v), v)$$

Proposition 4 *Suppose that b is symmetric, and consider an efficient market equilibrium with full consumer participation. Then, for every $w \in W$,*

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

Moreover, this expression decreases when the matrix $(\mu(\cdot|x))_{x\in X}$ undergoes Blackwell garbling.

Proof. Fix b, f, a. By definition, the price-per-draw of v is

$$PPD(v) = \frac{q \cdot \sum_w b(w|v)\pi(f(v), w)}{\sum_w b(w|v) \sum_y \pi(y, w)}$$

Also by definition,

$$\begin{aligned} \pi(x, w) &= \frac{\mu(x, w)a(x, w)}{\sum_{v\in f^{-1}(x)} b(w|v)} \\ \lambda_f(x, w) &= \frac{\sum_{v\in f^{-1}(x)} b(w|v)}{\sum_y \sum_{v\in f^{-1}(y)} b(w|v')} \end{aligned}$$

By assumption, $\mu(x, w) > 0$ implies $a(x, w) = 1$ and $\lambda_f(x, w) > 0$. Thus, whenever $\mu(x, w)\mu(y, w) > 0$, the above identities imply

$$\frac{\lambda(x, w)}{\lambda(y, w)} = \frac{\sum_{v\in f^{-1}(x)} b(w|v)}{\sum_{v\in f^{-1}(y)} b(w|v')} = \frac{\mu(x, w)}{\mu(y, w)} \cdot \frac{\pi(y, w)}{\pi(x, w)}$$

Efficiency implies (15). It follows that whenever $\mu(x, w)\mu(y, w) > 0$, we can write

$$\frac{\pi(y, w)}{\pi(x, w)} = \sqrt{\frac{\mu(y, w)}{\mu(x, w)}}$$

We can now plug the identities we have arrived at into the definition of $PPD(v)$, invoking the assumption that for any y, $b(w|v) = b(w|v')$ for all $v, v' \in f^{-1}(y)$:

$$
\begin{aligned}
PPD(v) &= \frac{q \cdot \sum_w \frac{\mu(f(v),w)}{n_f(f(v))}}{\sum_w \sum_y \frac{\mu(f(v),w)\pi(y,w)}{\pi(f(v),w)n_f(f(v))}} \\
&= \frac{q \cdot \mu(f(v))}{\sum_w \sum_y \mu(f(v),w)\frac{\pi(y,w)}{\pi(f(v),w)}} \\
&= \frac{q \cdot \mu(f(v))}{\sum_w \sum_y \mu(f(v),w)\sqrt{\frac{\mu(y,w)}{\mu(f(v),w)}}} \\
&= \frac{q \cdot \mu(f(v))}{\sum_w \sum_y \sqrt{\mu(f(v))\mu(y)\mu(w|f(v))\mu(w|y)}} \\
&= \frac{q}{\sum_y \sqrt{\frac{\mu(y)}{\mu(f(v))}} S(f(v),y)}
\end{aligned}
$$

By Remark 1, $PPD^*(w)$ decreases when $(\mu(\cdot|x))_{x \in X}$ is subjected to a Blackwell garbling. ■

Thus, when search costs are small, the equilibrium price–per-draw of keywords decreases as consumers' queries become less informative of their true wants.[2] To illustrate the comparative statics, suppose that $X = W$ and $\mu(x) = \frac{1}{|X|}$ for all x. Consider two extreme cases. First, suppose that $\mu(x|x) = 1$ for all x - i.e., consumers can perfectly describe their wants. Then, $PPD^*(w) = 1$ for every w under the efficient equilibrium induced by b^*. Second, suppose that $\mu(w|x) \approx \frac{1}{|X|}$ for every w, x (an exact equality would be inconsistent with the condition for an efficient equilibrium). In this case, there is virtually no correlation between consumers' favorite product and their query, and we have $PPD^*(w) \approx \frac{1}{|X|}$ for every w.

3.3 Efficient Equilibrium under $c = 0$

The maximal social welfare when $c = 0$ is 1, because every consumer type should end up getting the product he likes, and the duration of his search does not matter. Thus, a necessary and sufficient condition for (f, a) to induce an efficient outcome is that $\lambda_f(x, w) > 0$ and $a(x, w) = 1$ whenever $\mu(x, w) > 0$.

[2] If we dropped the full consumer participation assumption, such that $\lambda^*(x, w) = 0$ for some (x, w) in the support of μ, the formula for PPD^* would be somewhat messier, because the summation over w and y would not be universal.

Proposition 5 *There exists a broad match function b^* that induces an efficient market equilibrium (f, a) if and only if $\mu(\cdot|x) \neq \mu(\cdot|y)$ for every distinct x, y. In particular, we set f to be any onto function, and set b^* to be*

$$b^*(w|v) = \frac{\mu(f(v))\sqrt{\mu(w|f(v))}}{n_f(f(v))} \tag{24}$$

Proof. Any (f, a) that maximizes social welfare automatically satisfies condition (i) in Definition 1. In particular, note that f is onto, since $\mu(x) > 0$ for every x. The question is whether there exist such (f, a) that will also satisfy condition (ii).

Let us prove necessity first. Assume that $\mu(\cdot|x) = \mu(\cdot|y)$ for some distinct x, y. Inequalities (22) and (23) are necessary conditions for market equilibrium, because they are implied by condition (ii) in Definition 1. Rearranging these inequalities, and writing $\mu(w|x) = \mu(w|y) = \beta(w)$ for every w, we can see that the problem is to find a collection of real coefficients $(\psi(w))_{w \in W}$ such that

$$\begin{aligned} \sum_w \beta(w)\psi(w) &< 1 \\ \sum_w \frac{\beta(w)}{\psi(w)} &< 1 \end{aligned}$$

where

$$\psi(w) = \frac{\mu(x)\sum_{v \in f^{-1}(y)} b(w|v)}{\mu(y)\sum_{v \in f^{-1}(x)} b(w|v)}$$

To see why this is impossible, add the two inequalities:

$$\sum_w \beta(w)\left[\psi(w) + \frac{1}{\psi(w)}\right] < 2$$

But $\psi(w) + 1/\psi(w)$ attains a minimum at $\psi(w) = 1$, and since $\Sigma_w \beta(w) = 1$, we obtain a contradiction.

Let us turn to sufficiency. Suppose that $\mu(\cdot|x) \neq \mu(\cdot|y)$ for every distinct x, y. Fix some onto function f. By the definition of b^*, $b^*(w|v) = b^*(w|v')$ whenever $f(v) = f(v')$. It suffices to show that Inequality (21) can be satisfied. This inequality is simplified into

$$\frac{\mu(x)}{n_f(x)} > \sum_w \frac{b^*(w|v_x)\mu(y)\mu(w|y)}{n_f(y)b^*(w|v_y)}$$

whenever $f(v_x) = x$, $f(v_y) = y$. Now plug the definition of b^* as described in the statement of the proposition, and obtain the inequality

$$\sum_w \sqrt{\mu(w|x)\mu(w|y)} = \sqrt{S(x,y)} < 1$$

This inequality indeed holds whenever $\mu(\cdot|x) \neq \mu(\cdot|y)$. ■

The key argument in the proof of necessity is that the Bhattacharyya coefficient of two identical distributions ($\mu(\cdot|x) = \mu(\cdot|y)$) cannot be lower than one. The sufficiency argument exploits the property that $S(x,y) < 1$ whenever $\mu(\cdot|x) \neq \mu(\cdot|y)$, which implies a slack in the equilibrium requirement that each keyword is allocated to the firm type with the highest WTP. This slack gives us enough freedom in selecting an appropriate broad match function.

These arguments also reveal that if we did not require condition (ii) in Definition 1 to involve *strict* inequalities, it would be possible to construct a broad match function that implements an efficient equilibrium for all μ, simply by setting $b(w|v) = \mu(x)/n_f(x)$ whenever $f(v) = x$, such that $\psi(w) = 1$ for all w. This would imply that all firm types get a conversion rate of $1/|X|$ from all words. However, any slight perturbation in μ would upset condition (ii) in Definition 1, hence this equilibrium is not robust.

Finally, observe that when all products are equally popular (i.e., $\mu(x) = \frac{1}{|X|}$ for all x) and c is small, the necessary and sufficient conditions for market implementability of efficient outcomes under $c = 0$ and $c > 0$ coincide.

3.3.1 Example: Misinformation about Product Names

The following example illustrates the optimal broad match function under $c = 0$. Perhaps the simplest example of a gap between consumers' wants and their ability to describe them is when they are misinformed about the name of their desired product. In this case, broad matching can be viewed as a partial substitute for correcting misinformation. Let $X = W = \{x, y\}$. Assume

$$\begin{aligned}
\mu(x,x) &= \alpha(1-\varepsilon) \\
\mu(x,y) &= \alpha\varepsilon \\
\mu(y,y) &= (1-\alpha)(1-\varepsilon) \\
\mu(y,x) &= (1-\alpha)\varepsilon
\end{aligned}$$

where $\alpha > \frac{1}{2}$ and $\varepsilon < \frac{1}{2}$. The story is that the names of x and y are similar and thus easily confused with one another; α is the fraction of consumers who like product x; and ε is the (independent) probability that consumers are misinformed about product names.

The "rational expectations" aspect of condition (i) in Definition 1 means that the consumer's error cannot be interpreted as an accidental typo: the consumer type (x, y) genuinely believes that the name of the product he is looking for is y, and he does not reconsider this belief even after taking many unsuccessful draws from his search pool.[3] This is admittedly an extreme assumption, which shows the importance of extending the model to allow for consumer learning and multiple queries.

Under narrow matching, the equilibrium that maximizes social welfare is the following: for every $z \in X$, $f(z) = z$, $a(x, x) = a(y, y) = 1$, $a(x, y) = a(y, x) = 0$. That is, only consumers who know the correct name of their desired product engage in search, while the others give up on search. Social welfare is $1 - \varepsilon$. That is, only well-informed consumers are served. To take the opposite extreme, the following equilibrium maximizes social welfare under fully broad matching. Both words are allocated to the product x, and only consumer types who like x engage in search. Social welfare is α. That is, only consumers who like the popular product x are served. These extreme broad match functions illustrate the basic tension that our model captures. On one hand, narrow matching means that misinformed consumers are not served. Yet trying to resolve this market failure by fully broad matching causes a bigger market failure, whereby the product with mass appeal takes over the entire market and crowds out the "niche" product.

Let us now find an optimal broad match function. Because $c = 0$, efficiency requires that all consumer types engage in active search and eventually find their desired product. W.l.o.g, suppose that $f(z) = z$ for both $z = x, y$. If $b(z|z') > 0$ for all z, z', condition (i) in Definition 1 is satisfied. Let us turn to condition (ii):

$$\begin{aligned} b(x|x) \cdot \pi_{f,a}(x, x) + b(y|x) \cdot \pi_{f,a}(x, y) &> b(x|x) \cdot \pi_{f,a}(y, x) + b(y|x) \cdot \pi_{f,a}(y, y) \\ b(y|y) \cdot \pi_{f,a}(y, y) + b(x|y) \cdot \pi_{f,a}(y, x) &> b(y|y) \cdot \pi_{f,a}(x, y) + b(x|y) \cdot \pi_{f,a}(x, x) \end{aligned}$$

[3] To take a highly realistic example, think of an academic who Googles "Milgram" (respectively, "Milgrom") in search of material on a famous auction theorist (respectively, social psychologist).

Plugging the definition of π given by (12), we obtain

$$\begin{aligned}\alpha(1-\varepsilon)+\alpha\varepsilon &> \frac{b(x|x)(1-\alpha)\varepsilon}{b(x|y)}+\frac{b(y|x)(1-\alpha)(1-\varepsilon)}{b(y|y)} \\ (1-\alpha)(1-\varepsilon)+(1-\alpha)\varepsilon &> \frac{b(y|y)\alpha\varepsilon}{b(y|x)}+\frac{b(x|y)\alpha(1-\varepsilon)}{b(x|x)}\end{aligned}$$

Thus, we need to set two parameters:

$$\begin{aligned}r_1 &= \frac{b(x|x)(1-\alpha)}{b(x|y)\alpha} \\ r_2 &= \frac{b(y|x)(1-\alpha)}{b(y|y)\alpha}\end{aligned}$$

that satisfy the inequalities

$$\begin{aligned}1 &> r_1\varepsilon+r_2(1-\varepsilon) \\ 1 &> \frac{1}{r_1}(1-\varepsilon)+\frac{1}{r_2}\varepsilon\end{aligned}$$

Now, if we set

$$r_1=\frac{1}{r_2}=\sqrt{\frac{1-\varepsilon}{\varepsilon}}$$

both inequalities reduce to $\varepsilon(1-\varepsilon)<\frac{1}{4}$, which necessarily holds. Thus, any broad match function that satisfies the following equations sustains an efficient outcome in market equilibrium:

$$\begin{aligned}\frac{b(y|x)}{b(y|y)} &= \frac{\alpha}{1-\alpha}\sqrt{\frac{\varepsilon}{1-\varepsilon}} \\ \frac{b(x|y)}{b(x|x)} &= \frac{1-\alpha}{\alpha}\sqrt{\frac{\varepsilon}{1-\varepsilon}}\end{aligned}$$

It is easy to see that the function b^* given by (24) meets this requirement.

4 Mechanism Design

We return to the case of $c>0$. Our analysis so far established necessary and sufficient conditions for implementing the efficient collection of search pools $\lambda^*=(\lambda^*(x,w))_{x\in X,w\in W}$ as a competitive equilibrium of the market for keywords. This raises the question of whether the first-best can be achieved under weaker conditions, if the allocation of firms to consumers' search pools is done via some general incentivecompatible *mechanism*. We conceive of a mechanism as a game form in which the players are the *firms*; consumers make their individually rational search decisions given the firms' equilibrium behavior in the mechanism-induced game.

More specifically, we consider direct anonymous mechanisms for allocating firms to search pools. Each firm reports a type $\hat{x} \in X$. Given the profile of reports, each firm is assigned a pair (T, p), where T is a monetary transfer to the intermediary, and $p \in \Delta(2^W)$ is a probability distribution over subsets of W, where $p(V)$ is the probability that the set of search pools to which the firm gets access is V. The restriction to direct mechanism follows from the revelation principle; anonymity means that the mechanism treats identically firms that submit the same report. This means that when all firms report truthfully, a certain fraction of the x firms, denoted $p_x(V)$, enters each of the search pools in V, and each x firm pays a certain transfer denoted T_x.

Given a truth-telling Nash equilibrium, we can define the probability that a firm reporting x enters the search pool associated with a given w :

$$q(x, w) \equiv \sum_{V \subseteq W | w \in V} p_x(V) \tag{25}$$

The probability that a consumer of type (x, w) will find a firm x in the pool w is

$$\lambda(x, w) = \frac{q(x, w)}{\sum_y q(y, w)} \tag{26}$$

It follows that for every pair of products x and y that are assigned to the search pool w,

$$\frac{\lambda(x, w)}{\lambda(y, w)} = \frac{q(x, w)}{q(y, w)}$$

The problem facing the mechanism designer thus boils down to choosing a collection $(T_x, p_x)_{x \in X}$ that maximizes social welfare as defined in (13), where the dependence of $\lambda(x, w)$ on p_x is given by (25) and (26), subject to the incentive-compatibility constraint that for every $x, y \in X$,

$$\sum_{V \subseteq W} p_x(V) \cdot \sum_{w \in V} \frac{\mu(x, w) a(x, w)}{\sum_{V \subseteq W | w \in V} p_x(V)} - T_x > \sum_{V \subseteq W} p_y(V) \cdot \sum_{w \in V} \frac{\mu(x, w) a(x, w)}{\sum_{V \subseteq W | w \in V} p_x(V)} - T_y \tag{27}$$

where a satisfies condition (i) in Definition 1 - that is, consumers choose to search if and only if their expected search cost is below 1. Note that we adopt here a strict inequality, as in our definition of market equilibrium. Since we do not impose any constraints on the mechanism designer's budget at this stage, we can ignore the firms' participation constraint.

Lemma 3 *A mechanism defined by* $(T_x, p_x)_{x \in X}$ *implements* λ^* *in truth-telling (strict) Nash equilibrium if and only if*

$$\mu(x) - \sqrt{\mu(x)\mu(y)S(x,y)} > T_x - T_y \tag{28}$$

for every distinct x, y *for which* $\lambda^*(x,w)\lambda^*(y,w) > 0$ *for some* w.

Proof. We begin by rewriting the incentive constraint (27) in terms of $q(x,w)$:

$$\sum_{w \in W} q(x,w) \cdot \frac{\mu(x,w)a(x,w)}{q(x,w)} - T_x > \sum_{w \in W} q(y,w) \cdot \frac{\mu(x,w)a(x,w)}{q(x,w)} - T_y \tag{29}$$

First, note that for any x, w for which $\lambda^*(x,w) = 0$, we can design p_x such that $q(x,w) = 0$, and so consumer type (x,w) will choose $a(x,w) = 0$. If $\lambda^*(x,w) > 0$, then it must be the case that $a(x,w) = 1$ in equilibrium, following the same argument as in Section 3. It follows that w.l.o.g, we can ignore the term $a(x,w)$ in inequality (29), which is thus reduced to

$$\mu(x) > T_x - T_y + \sum_w \frac{q(y,w)}{q(x,w)}\mu(x,w) = T_x - T_y + \sum_w \frac{\lambda(y,w)}{\lambda(x,w)}\mu(x,w) \tag{30}$$

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 λ^* satisfies (30) if and only if

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

which completes the proof. ■

The next result establishes that whenever different preference types have different conditional query distributions, the first-best is implementable by some anonymous direct mechanism.

Proposition 6 *Suppose that $\mu(\cdot|x) \neq \mu(\cdot|y)$ for every distinct x, y. Then, there is an anonymous direct mechanism that implements the efficient outcome.*

Proof. Fix some x, y for which $\lambda^*(x,w)\lambda^*(y,w) > 0$ for some w. By Lemma 3, the relevant IC constraint that prevents type x from pretending to be y, denoted $IC(x,y)$, is given by the inequality,

$$\mu(x) - \sqrt{\mu(x)\mu(y)S(x,y)} > T_x - T_y \tag{31}$$

Let $\theta(x,y)$ denote the L.H.S. of (31), and rewrite the constraint $IC(x,y)$ as $\theta(x,y) > T_x - T_y$. Since $\mu(\cdot|x) \neq \mu(\cdot|y)$, we have $S(x,y) < 1$, and hence, for any cycle of products $(x_1, x_2, \ldots x_m, x_1)$,

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

Consider a complete weighted directed graph, whose set of nodes is X, and the weight on the link from the node x to the node y is $\theta^*(x,y) = \theta(x,y) - \varepsilon$, where $\varepsilon > 0$ is sufficiently small such that the sum of weights along any cycle is strictly positive. Define the weight of a link from x to itself as $\theta^*(x,x) = 0$. A *path* from x to y is a sequence of nodes that begin with x and end 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 strictly positive, the distance is always well-defined and it satisfies the triangle inequality: for any x, y, z, $\delta(x,z) \leq \delta(x,y) + \delta(y,z)$.

Now, fix some $x^* \in X$. For any $x \in X$, define $T_x = \delta(x, x^*)$.[4] By the triangle inequality,

$$\theta^*(x,y) + \delta(y,x^*) \geq \delta(x,x^*)$$

[4] We thank Yossi Azar for suggesting the shortest-path method for finding T

for any x, y. Since $\theta^*(x,y) < \theta(x,y)$, this implies that for any pair of distinct products x, y, $\theta(x,y) > T_x - T_y$, which is equivalent to $IC(x,y)$. ■

The proof of Proposition 6 relies on a formal analogy to the problem of finding shortest paths in a weighted directed graph. The set of nodes in the graph is X, and every weighted link represents a potential IC constraint: the weight on the link from x to y is the gross utility loss for an x firm pretending to be a y firm (this loss can be negative, of course). The structure of λ^* implies that the sum of these weights along any cycle is strictly positive. This means that the notion of a distance from one node to another is well-defined and satisfies the triangle inequality (see Cormen et. al. (2001)). If we define the transfer T_x as the distance from x to some reference node x^*, the IC constraints are essentially restatements of triangle inequalities.

Thus, as long as there are no restrictions on the transfers that the SI can administer, the first-best is implementable for generic consumer type distributions. However, note that the competitive market for keywords had the additional feature that firms' surplus was fully extracted. This raises the following question. Suppose that we are interested in both efficiency and full surplus extraction; can we design an anonymous direct mechanism that implements these twin objectives when the market mechanism fails? The answer turns out to be negative.

Proposition 7 *There exists an anonymous direct mechanism that induces an efficient Nash equilibrium in which firms earn zero profits, if and only if there exists a broad match function that induces an efficient equilibrium in the market for keywords.*

Proof. Consider the constraint $IC(x,y)$, as given by (28). In order for x firms and y firms to earn zero profits in equilibrium, we must have $\mu(x) - T_x = \mu(y) - T_y = 0$. Thus, $IC(x,y)$ and $IC(y,x)$ are reduced to

$$\begin{aligned} \mu(y) &> \sqrt{\mu(x)\mu(y)S(x,y)} \\ \mu(x) &> \sqrt{\mu(x)\mu(y)S(x,y)} \end{aligned}$$

which is equivalent to the condition that

$$\frac{\mu(x)}{\mu(y)}S(x,y) < 1$$

for both x, y. This is precisely the necessary and sufficient condition for implementing the first-best constraint in market equilibrium. ■

To see why the latter result is not obvious a priori, note that in general, the IC constraint (27) does *not* coincide with condition (*ii*) in Definition 1, even if we ignore the transfers. First, the conditions superficially *look* different. The former requires that x firms prefer the distribution p_x over collections of search pools to the distribution p_y, for any $y \neq x$. Thus, *both* sides of the constraint show how an x firm evaluates distributions over collections of search pools. In contrast, the corresponding condition for market equilibrium requires that the value of a keyword v to a firm of type $f(v)$ be strictly higher than the value of that keyword to any other firm type. Thus, each side of the constraint displays the value of a keyword to a *different* firm type.

The two conditions can be written in ways that clarify their essential difference. As the proofs of Lemma 3 and Proposition 7 show, the IC constraint in the direct mechanism (coupled with the zero-profit requirement) can be written as follows:

$$\mu(x) > \sum_w \frac{\lambda(y,w)}{\lambda(x,w)} \mu(x,w)$$

In contrast, if we sum the inequalities given by condition (*ii*) of Definition 1 over all words $v \in f^{-1}(x)$, and then plug (10), we obtain the following necessary condition:

$$\mu(x) > \sum_w \frac{\lambda(x,w)}{\lambda(y,w)} \mu(y,w)$$

The R.H.S of these two inequalities are clearly different. However, the first-order conditions that characterize λ^* - specifically, the key identity (15) - imply that *at the first-best*, the R.H.S of the two conditions do coincide. Whether this coincidence has a deeper significance is an interesting question for future research.

5 Ordered search

In this section we consider a competitive market for keywords in which the SI 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 SI now *optimally chooses* which firm type the consumer will encounter at each draw, as a function of his search history.

An ideal centralized SI would use its direct knowledge of firms' types to fix the exact order. The sequence of product types that maximizes the total surplus of consumers who submit a given query w is determined according to a simple maximum-likelihood calculation. For expositional simplicity, suppose that c is sufficiently low, such that efficiency would require full consumer participation.

The first product type to be displayed, denoted $x_1(w)$, is most likely to be the consumer's favorite product conditional on his vocabulary 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 $x_k(w)$, will be the product type that is 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 q tends to zero, the k-th product type on the list will simply be the k-th most popular product among w consumers.)

In contrast to this omniscient SI, our market-based SI will determine the sequence according to the firms' equilibrium market behavior - namely, the keywords they choose to pay for. We need to adapt the notion of broad matching to environments with ordered search. For every $w, v \in W$, $b(w|v)$ denotes a *probability distribution over positions* $1, 2, 3, ...$ in the search pool associated with the query w. That is, $b_k(w|v)$ is the probability with which a firm that pays for v gets access to the k-th position on the list that is displayed to a consumer who submits the query w. The firm that eventually appears in this position will be randomly drawn from the collection of all the firms that were granted access to it via the broad match function. 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. Armed with this extended definition of b, we can calculate the number of transactions that each firm expects when it pays for a word, and this is the firm's WTP for the word. The definition of market equilibrium can be extended accordingly: f should allocate each word to the firm type with the highest WTP.

Consider the following broad match function. Let f be an arbitrary onto function. For every $w, v \in V$:

$$b_k(w|v) = \begin{cases} \frac{1}{n_f(f(v))} & if \quad f(v) = x_k(w) \\ 0 & if \quad f(v) \neq x_k(w) \end{cases}$$

Thus, the k-th position on the list of a w consumer will be randomly allocated among all the firms that pay for the words that are allocated to the firm type $x_k(w)$.

Let us now show that under this broad match function, f is consistent with market equilibrium (recall that we assume low search costs, such that full consumer participation is consistent with both efficiency and individual consumer rationality). Suppose that an x firm considers paying for a word that f allocates to y firms. This will give him 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 these positions. Moreover, the n_f-normalization in the definition of b ensures that the total measure of firms that get access to any position on any consumer's list is 1. It follows that x firms will have a lower WTP for the positions that are meant for y firms, in accordance with the requirement of market equilibrium.

Note that from a mechanism-design point of view, there is a natural indirect mechanism that implements the efficient outcome in truth-telling Nash equilibrium: for each keyword independently, the SI can sequentially auction off each position on the list.

The lesson from this section is perhaps that the problem of implementing an efficient search environment as an equilibrium outcome in a "market for keywords" is somewhat trivial when the SI can *fully* determine the order in which consumers inspect alternatives. It appears that interest in broad matching in markets for search platforms arises when the SI has only an *imperfect* ability to control the order of inspection (for reasons that were listed in Section 2 - see the comment on search technology).

6 Profit Maximization

Throughout this paper, we assumed that the SI benevolently maximizes social surplus, because our objective was to explore the welfare properties of a competitive market for search platforms when consumers provide noisy signals of their preferences. How would our analysis change if we assumed alternatively that the SI maximizes its profits?

First, consider the case of small c, where the socially optimal collection of search pools λ^* has the feature that $\lambda^*(x, w) > 0$ whenever $\mu(x, w) > 0$. If λ^* can be sustained in market equilibrium, then every consumer ends up transacting. Since firms surrender their entire surplus to the SI in market equilibrium, the SI's total profit is 1, which is as high as it can get. It follows that if λ^* implies full consumer participation and can be sustained in market equilibrium, our analysis is consistent with maximization of the search engine's profit. When c is large, there may be a conflict between maximizing welfare and maximizing the SI's profit, because the latter does not take negative search externalities into account. It is easy to construct examples in which full consumer participation will be consistent with market equilibrium but not with maximizing social welfare. A profit maximizing SI would opt for the former.

The observation that a monopolistic profit-maximizing SI may have an incentive to degrade the quality of consumer search has been made in the literature in various contexts (see, for example, Eliaz and Spiegler (2011a) and Hagiu and Jullien (2011)). The source of this tension in the present paper is that consumers with different preferences may share the same signal, which generates negative search externalities among them that a profit-maximizing search engine neglects. We are not aware of previous papers that addressed this particular source.

7 Conclusion

This paper addressed the following general question: *under what conditions is a decentralized competitive market efficient in helping individuals find objects they need, where the objects of trade in the market are noisy signals of the individuals' needs?* Our leading example considered consumers who search for products by submitting queries that only partially describe what they are looking for. In this context, our question could be rephrased as follows: *suppose that a benevolent search intermediary switches from a centralized matching algorithm to a pure market system of "sponsored search"; 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 assigning objects to search pools. For example, online recommender systems assign search pools to individuals according to their past behavior (including purchases, web browsing, search history and mail content), which serves as an imperfect signal of the individuals' 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 recommended articles.

To see how our model 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 probabilistic access to some set of profiles. In this context, our question can be rephrased as follows: *suppose that a recommender system such as Netflix abandons its centralized recommendation algorithm in favor of a "market for sponrored recommendations"; will the quality of* *its recommendations deteriorate as a result?* Our main result suggests 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).[5]

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)).[6] 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". However, the topics are latent and implicit (unlike our model, where W is the set of keywords *explicitly* used by consumers); 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 latest topics (whose size is fixed a priori; unlike our model, here $|W| \ll |X|$). After estimating the distribution, the recommender system often applies the Bhattacharyya/Hellinger measure of similarity in order 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' market incentives.

References

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

[2] Armstrong, M. "Competition in two-sided markets," RAND Journal of Economics 37.3 (2006): 668-691.

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

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

[3] Athey, S. and G. Ellison (2011), "Position Auctions with Consumer Search," The Quarterly Journal of Economics, Oxford University Press, vol. 126(3), pages 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 & egg: Competition among intermediation service providers." RAND journal of Economics: 309-328.

[11] Chen, Y., C. He (2011), "Paid Placement: Advertising and Search on the Internet," Economic Journal, Royal Economic Society, vol. 121(556), pages F309-F328, November.

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

[13] Dhangwatnotai, P. (2011), "Multi-keyword sponsored search," Proceedings of the 12th ACM conference on Electronic commerce.

[14] 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.

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

[16] Eliaz, K. and R. Spiegler (2011b), "On the Strategic Use of Attention Grabbers," Theoretical Economics 6, 127-155.

[17] Hagiu, A. and B. Jullien (2011), "Why do intermediaries divert search?," RAND Journal of Economics, RAND Corporation, vol. 42(2), pages 337-362, 06.

[18] Kranton, R. and D. Minehart (2001), "A Theory of Buyer-Seller Networks," American Economic Review, American Economic Association, vol. 91(3), pages 485-508, June.

[19] Rochet, J. and J. Tirole (2003), "Platform Competition in Two-Sided Markets," Journal of the European Economic Association, MIT Press, vol. 1(4), pages 990-1029, 06.

[20] Spiegler, R. (2000), "Extracting Interaction-Created Surplus," Games and Economic Behavior, Elsevier, vol. 30(1), pages 142-162, January.

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