Joint Discussion Paper
Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 48-2014

Tim Brühn, Georg Götz and Annette Meinusch

The Value of User-Specific Information for Two-Sided Matchmakers

Coordination: Bernd Hayo • Philipps-University Marburg
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

The Value of User-Specific Information for Two-Sided Matchmakers

Tim Brühn*, Georg Götz,**

and Annette Meinusch***

This article analyzes the incentives of a monopolistic matchmaker to generate user-specific information. By merging two-sided market modeling with two-sided matching, we derive a micro-foundation of cross-side externalities as a function of the number of potential matches and the accuracy level of user-specific information. Incentives to make fixed investments in identification technologies are determined by two effects that work in opposing directions: Whereas economies of scale work in favor of platforms with large customer bases, expected improvements to match quality are more significant for small-scale platforms.

* *Tim.Bruehn@wirtschaft.uni-giessen.de, Justus-Liebig-Universität Gießen*

** *Georg.Goetz@ wirtschaft.uni-giessen.de, Justus-Liebig-Universität Gießen*

*** *Annette.Meinusch@wirtschaft.uni-giessen.de, Justus-Liebig-Universität Gießen*

This article was presented at the 41st Annual Conference of the European Association for Research in Economics in Milan, August 2014. We are grateful to Zhijun Chen, Toker Doganoglu, Ali Hortaçsu, Tobias Klein, Robin Lee, Martin Peitz, Thomas Wagner, and Tobias Wolkober for discussions.

1. Introduction

■ We leave digital footprints any time we browse the web. Platforms track them to learn about our characteristics and preferences. This requires significant investment in identification technology on the part of platform providers, such as advertising networks, social media, online dating services, and search engines. Google, for one, goes to great lengths to identify user characteristics by tracking individual browsing patterns through cookies and indexing web contents by means of software robots, web crawlers, and spiders. The firm's capital expenditures to back up these activities were some seven billion dollars in 2013.[1] Dating services deploy elaborate personality tests and incentivize agents to self-select into specific categories of characteristics during the subscription process. Facebook and many other platforms continuously glean information about their users' preferences: The 'like' button, the recommendations feed, and other fancy technologies ultimately serve to create high-quality personal data or user profiles.

The reason for such massive investments in identification technology is straightforward enough: By increasing the accuracy of agent-specific information, platforms increase the quality of intermediation between heterogeneous match partners. Specifying the platform operator's optimal level of investment and its determinants, however, requires more detailed study. The present article thus analyzes incentives to identify user-specific information from the perspective of a monopolistic matchmaker in a two-sided market along the following lines. The matchmaker coordinates the allocation between members of two groups of

---

[1] http://www.datacenterknowledge.com/archives/2014/02/03/google-spent-7-3-billion-data-centers-2013/

horizontally differentiated agents (e.g., users and advertisers). We assume that an advertiser values matches based on the fit between the brand positioning and the preferences of the user targeted by the advertisement. The ideal user is the one whose preference coincides perfectly with the advertising message. Advertisers thus have assortative preferences for match partners (Hitsch, Hortaçsu, and Ariely, 2010; Kalmijn, 1998; Shimer and Smith, 2000). The more significant the mismatch between the advertising message and the user's preference, the lower the utility advertisers receive from matching.

Generating information about user-specific preferences and characteristics enables the platform operator to assign agents to different categories, such as demographic characteristics, keywords, or tags. The more categories the platform operator is able to distinguish, the more accurate the agent-specific information. Based on this information, the platform operator coordinates advertisements to users' eyeballs in a (complete) one-to-one matching that involves all agents active on the platform.

In the first part of this article, we derive an efficient matching mechanism for given samples of users and advertisers. This mechanism is required because users with random preferences join the platform, thus making imperfect matches possible. We then proceed to derive the probability distribution for corresponding match qualities in a closed form. The distribution of match qualities not only determines the cross-side externalities received by advertisers as a function of the number of platform users and the accuracy of identification. It also provides a micro-foundation for the interaction parameter implemented in the canonical models of Armstrong (2006), and Rochet and Tirole (2003, 2006).

In the second part of the article, we analyze investment incentives as a function of market size (i.e., an exogenously given number of users and advertisers) and find two opposing effects. First, information costs per match decrease with market size due to economies of scale resulting from what are typically fixed investments in identification technology. Mark-ups per match, and thus investment incentives, increase with market size. Second, and in the opposing direction, more accurate identification has a positive effect on match quality, and this effect decreases with market size. Assuming that the platform operator provides a service to users and advertisers with a given range of preferences and characteristics, the subscription of users with random preferences coincides with a relatively high probability for imperfect matches in small markets and a rather low probability for imperfect matches in large markets. Comparing the assignments of advertisements to users before and after making a marginal investment, additional agent-specific information leads to a more efficient (re)allocation of pairs that were seen as optimally matched before. Loosely speaking, the most significant effect of this reallocation on match quality is to avoid matches of the worst quality. As the probability for these matches is higher in small markets, the marginal effect of generating additional information is higher for platforms with a small rather than a large customer base.

Our article combines two-sided matching with two-sided market modeling. To the best of our knowledge, it is the first analysis of a platform operator’s incentives to invest in agent-specific information and match quality. This topic is potentially relevant for social media and location-based advertising, online dating services, real-estate brokerage, M&A services in investment banking.

The article is organized as follows: The relevant literature is reviewed in Section 2. Section 3 introduces the ingredients of the model: timing of the game, information generating process, and utility functions. Section 4 discusses the matching and derives an efficient matching mechanism, the probability distribution of match qualities, and inconvenience costs. Section 5 analyzes incentives to invest in generating information as a function of market size. Section 6 concludes.

2. Related literature

■ Studying the impact of agent-specific information on match quality and profits relates to three strands of literature. The first of these is two-sided market literature. The seminal articles in this field deal with the profit-maximizing pricing strategies of platform operators (Armstrong, 2006; Parker and Van Alstyne, 2005; Rochet and Tirole, 2003, 2006; Weyl, 2010). The price level and price structure are determined by the (relative) strength of cross-side externalities. They determine the utility of agents with complementary businesses and are a function of the potential and total number of partners with whom a given agent can interact. Interaction partners are typically assumed to be homogeneous in a horizontal dimension (preferences and characteristics) and in a vertical dimension (quality).[2] On many platforms controlled by an intermediary, however, subscribers typically interact only with a subset of heterogeneous match partners, and maybe with just one partner. In this case, the strength of cross-side externalities may be interpreted as the probability of being involved in a high-quality match. We contribute to this line of research by deriving a micro-foundation for this definition of cross-side externalities.

---

[2] Heterogeneity is introduced in how agents value membership (Armstrong, 2006), interactions (Rochet and Tirole, 2003), or both (Weyl, 2010).

The second related strand of literature deals with efficient two-sided matching. Our platform operator engages in efficient matching, which maximizes the total surplus of the horizontally differentiated agents involved. Efficient matching is particularly important if a platform operator provides intermediation services in a market characterized by positive intertemporal demand externalities, such that the actual match quality affects the likelihood of subscription in future periods.[3] Kuhn (1955) and Munkres (1957) solve the efficient matching problem using the so-called 'Hungarian Algorithm'. This algorithm produces a global value-maximizing one-to-one matching given a matrix of agent-specific valuations for each potential match. Demange, Gale, and Sotomayor (1986) reformulate the assignment problem as a multi-item auction.[4]

Both of these approaches are gradual ones and work under general conditions. We assume that agents have assortative preferences for match partners, which implies a clear structure of valuations for potential matches. Given this structure, we are able to implement a less complex mechanism to generate an efficient matching. Our match mechanism resembles histogram-matching approaches typically used in computer science (Shen and Wong, 1983) and enables us to calculate expected match values in a closed form.

Third, this article is related to literature on information acquisition in two-sided matching markets. Lee and Schwarz (2012) analyze the incentives of agents (firms)

[3] Damiano and Lee (2007), whose model is discussed in more detail below, also show that efficient one-to-one matching is in the interest of matchmakers in that it maximizes the total value received by agents, and thus matchmakers' profits.
[4] The auction consists of a number of stages. At each step, the auctioneer identifies sets of agents that are overdemanded. She increases prices for each matching involving overdemanded agents, thus gradually sorting out inefficient couples until a global value-maximizing allocation results.

to acquire information about potential match partners (workers), and the authors derive the optimal level of investment into information acquisition (costs resulting from the number of job interviews conducted). By contrast, we focus on the optimal investments to be made by an intermediary who seeks the profit-maximizing level of match quality.

Damiano and Li's (2007) model is perhaps closest to our work. The authors set up a two-sided market model with agents who are vertically differentiated and who have homogeneous preferences (the higher, the better) for the quality of their respective match partner. A monopolistic platform operator, who ex ante faces users with known preferences, installs a number of meeting places (categories) where potential match partners are matched randomly. The operator announces a subscription price for each meeting place to incentivize agents to self-select into these categories. Given all this, there is efficient matching by design. We also model a two-sided market, but deal with agents who are horizontally differentiated and who have assortative preferences for their respective match partner (i.e., preferences are a function of the type of agent). Instead of assuming deterministic subscription of agents, we assume that the platform operator faces users with random preferences. Thus, imperfect matches are possible and have to be coordinated by a matching mechanism.

3. The model

■ The timing of the game and the decisions agents take are as follows: In the first stage, the platform operator sets prices for placing advertisements in each category and decides about the accuracy of identification, i.e., the number of categories. In the second stage, advertisers and users subscribe to the platform service. Agents subscribe whenever they receive a non-negative utility from doing so. The decision of advertisers is determined by two parameters: the number of platform users and the expected accuracy of the operator's identification technology. As is shown below, both factors increase the probability of high-quality matching. In the third stage, the platform operator matches advertisements with users.

To solve the model, we first need to specify the information generating process and the utility functions. We then proceed in a recursive way to the timing of the game: We derive an optimal matching mechanism and the resulting structure of the assignment employed by the platform operator. Given the structure of the assignment, we calculate agents' willingness to pay. We are then able to set up the profit function and derive both the profit-maximizing prices and the incentives to invest in the identification technology.

□ **Information generating process.** Potential subscribers to the platform service are advertisers A and users U. Without loss of generality, we assume that the number of advertisers and users is the same and equal to market size M. A single user i on market side U has an inherent preference x_i^U. An advertiser j has a brand positioning x_j^A. Preferences and brand positionings are uniformly distributed over a unit Hotelling interval (Hotelling, 1929). Without investing in identification technology, the platform operator does not generate agent-specific information. The preference of a user i is a random variable X^U with realizations x^U distributed over the line segment $\Psi \coloneqq [0,1]$. Similarly, the brand positioning x_j^A of a single brand j is a random variable X^A with realizations x^A distributed over $\Psi \coloneqq [0,1]$.

FIGURE 1

If the platform operator invests in identification technology and generates agent-specific information, the operator increases the accuracy of agent-specific information. The accuracy level is captured by the number of categories k to which the operator can assign agents. Each category is defined by a specific category number a and u, with $a, u = \{1, \dots, k\}$. In Figure 1, categories are depicted by the subintervals ψ_u^U and ψ_a^A. The unit interval on both market sides is segmented into k subintervals of equal length $\psi_u^U \coloneqq \left[\left(\frac{u-1}{k}\right), \left(\frac{u}{k}\right)\right] \epsilon\, \Psi$ and $\psi_a^A \coloneqq \left[\left(\frac{a-1}{k}\right), \left(\frac{a}{k}\right)\right] \epsilon\, \Psi$. The upper bound of each subinterval is given by $\left(\frac{u}{k}\right)$ on the user side and $\left(\frac{a}{k}\right)$ on the advertiser side. The lower bounds are $\left(\frac{u-1}{k}\right)$ and $\left(\frac{a-1}{k}\right)$.

The user-specific preference X^U or the advertiser's brand position X^A that fall into a subinterval are transformed into category-specific random variables X_u^U and X_a^A with realizations x_u^U and x_a^A distributed uniformly over the corresponding subintervals $\psi_u^U \epsilon\, \Psi$ and $\psi_a^A \epsilon\, \Psi$. The probability density functions are given by $f(x_u^U) = \frac{1}{\left(\frac{u}{k}\right) - \left(\frac{u-1}{k}\right)} = k$ and $f(x_a^A) = \frac{1}{\left(\frac{a}{k}\right) - \left(\frac{a-1}{k}\right)} = k$.

□ **Utility functions.** In defining the utility of agents, we make simplifying assumptions on users' utility to focus on the advertiser's side.

First, we assume that users subscribe for free. This assumption captures internet platforms that mainly generate revenue via advertising, such as search engines, online video platforms and social networks (Evans and Schmalensee, 2007). Second, we assume that a user's stand-alone value received from using the platform service is always higher than the inconvenience costs she may incurs from advertisements.

A user i with preference x_i^U receives a utility $v(x_i^U)$ given by

$$v(x_i^U) = V^U - t^U(x_j^A - x_i^U)^2.$$

User i has a reservation value for subscription V^U and incurs inconvenience costs $t^U(x_j^A - x_i^U)^2$ from advertising that are weakly lower than V^U, such that $t^U \leq V^U$. The level of sensitivity to advertisements is captured by t^U, which is multiplied by the individual squared distance between the inherent preference x_i^U of user i and the characteristic x_j^A of the matched advertisement j. Given our assumptions, subscription takes place even if an (expected) advertisement x_j^A and the user-specific preference x_i^U are positioned on opposite extremes of Ψ.

An advertiser's utility is determined by the platform operator, as the platform operator allocates advertisements to users. Advertisers' utility is then determined by the quality of a one-to-one match of an advertisement x_j^A with a user who prefers x_i^U.

Advertiser j's utility $v(x_j^A)$ reads as

$$v(x_j^A) = V^A - p^A - \theta^A(x_j^A - x_i^U)^2.$$

The equation says that utility is determined by a reservation value V^A for an ideal match, inconvenience costs $\theta^A(x_j^A - x_i^U)^2$, and the price p^A paid per match. As explained in the introduction, imperfect matches yield lower utility. We account for a possible misfit between the advertising content x_j^A and the preference of the user x_i^U by allowing for inconvenience costs. We assume that inconvenience costs increase more than proportionally with distance. The term $(x_j^A - x_i^U)^2$ denotes the squared distance between matching partners x_j^A and x_i^U. The parameter θ^A captures the sensitivity to mismatches.

However, in the moment an advertiser decides whether to place an advertisement or not, she does not know what kind of user she is matched with. Given the operator's accuracy level of user-specific information, an advertiser faces two sources of uncertainty. First, there is uncertainty about the category u into which the preference of match partner i falls. Second, there is uncertainty about the exact position x_i^U of a user's preference within that certain category u.

Assuming we know the category of a user i, we start by deriving the second source of uncertainty. As an advertiser knows both the position of her advertisement on the Hotelling line x_j^A and the fact that the preference of an expected match partner i is distributed in the interval $\psi_u^U := \left[\left(\frac{u-1}{k}\right), \left(\frac{u}{k}\right)\right]$ with the probability density $f(x_u^U)$, we derive expected inconvenience costs by integrating user i's expected position over the subinterval ψ_u^U. Inconvenience costs incurred from a match between a specific advertiser positioned on x_j^A and a user in category u are therefore given by

$$E^A\left[\theta^A\left(x_j^A - x_i^U\right)^2 | u\right] = \theta^A \int_{\frac{u-1}{k}}^{\frac{u}{k}} \left(x_j^A - x_u^U\right)^2 f(x_u^U)\, \mathrm{d}\, x_u^U.$$

Given inconvenience costs for a specific match between an advertiser located on x_j^A and a user in category u, we are able to model the first uncertainty about mismatches between advertisers in category a and users in category u. We assume that the platform service is used by users who have different (random) preferences. Thus, preferences x_i^U are identical and independently distributed (i.i.d.) random variables drawn from equally distributed preferences in the population. If this is the case, there is a positive probability for imperfect matches $(a \neq u)$. An advertiser's inconvenience costs are based on probabilities for matches of different qualities. We denote the probability for a match between an advertisement of category a and a user of category u by $w(a, u)$.

For illustration, see Figure 2: If there is an advertiser located in category $a = 1$, $w(1,1)$ denotes the probability of this advertiser being matched to a user located in category $u = 1$. $w(1,2)$ refers to the probability for a match with a user of category $u = 2$, and $w(1,3)$ denotes the probability for a match between $a = 1$ and $u = 3$.

FIGURE 2

The expected utility an advertiser on x_j^A receives from subscription is then determined by V^A and p^A and the weighted sum of all possible category-specific inconvenience costs

$$E^{A}\left(v(x_j^A)\right) = V^{A} - p^{A} - \theta^{A} \sum_{u=1}^{k} w(a,u) \int_{\frac{u-1}{k}}^{\frac{u}{k}} (x_j^A - x_u^U)^2 f(x_u^U)\, \mathrm{d}\, x_u^U. \qquad (1)$$

As derived above, match probabilities $w(a,u)$ are central in our model. They determine the expected utility of advertisers and their willingness to pay. In the following subsections, we show how match probabilities $w(a,u)$ are derived from the matching mechanism the platform operator employs.

4. Matching

■ We now change from the perspective of the advertiser (A) to that of the platform operator (P) and discuss the matching mechanism and its implications for the advertiser's willingness to pay for intermediation. We require the matching mechanism to have the following properties:

1. The matching is complete, i.e., every advertisement j and user i is involved in the matching.
2. Advertisements and users are matched one-to-one, i.e., the platform operator sends exactly one advertisement to one user.
3. Advertisers do not act strategically, i.e., advertisements coincide perfectly with brand positioning.
4. The matching is globally efficient given the quality of agent-specific information. A matching is efficient if the total inconvenience costs of all advertisers and users involved in the matching are minimized.

In the following subsection, we describe how the matching mechanism is determined by the platform operator's accuracy level of agent-specific information. Subsequently, we derive an efficient assignment of advertisements to users given incomplete agent-specific information. The structure of the assignment determines match probabilities $w(a,u)$ that advertisers take into account when deciding about subscription.

□ **Inconvenience costs observed by the platform operator.** Contrary to advertisers and users, the platform operator (P) holds information about both advertisements and preferences only on the interval levels ψ_a^A and ψ_u^U. This means that the platform operator only knows each advertiser's and each user's category number u and a and the probability distributions $f(x_a^A)$ and $f(x_u^U)$ of their respective characteristic and preference in the category interval ψ_u^U and ψ_a^A. From the perspective of the platform operator (P), the expected inconvenience costs $E^P[ic(a,u)]$ of an advertiser of category a incurred by a match with a user of category u read as

$$E^P[ic(a,u)] = \theta^A \int_{\frac{a-1}{k}}^{\frac{a}{k}} \left(\int_{\frac{u-1}{k}}^{\frac{u}{k}} (x_a^A - x_u^U)^2 \, f(x_u^U) \, \mathrm{d}\, x_u^U \right) f(x_a^A) \, \mathrm{d}\, x_a^A. \qquad (2)$$

Not knowing the exact position of either the advertisement j or the preference i on their category intervals, expected inconvenience costs are calculated with a double integral, see Equation 2. The inner integral shows inconvenience costs for a specific advertiser j on a hypothetical position $x_{j,a}^A$ being matched to a user with preference i

located in the subinterval $\psi_u^U \coloneqq \left[\left(\frac{u-1}{k}\right), \left(\frac{u}{k}\right)\right]$. The outer integral lets this hypothetical position $x_{j,a}^A$ vary within the boundaries of the subinterval $\psi_a^A \coloneqq \left[\left(\frac{a-1}{k}\right), \left(\frac{a}{k}\right)\right]$. Under the assumption of uniformly distributed preferences $x_{i,u}^U$ and advertisements $x_{j,a}^A$, Equation 2 simplifies to

$$E^P[ic(a,u)] = \frac{\theta^A}{6\,k^2} + \frac{\theta^A}{k^2}(a-u)^2 \,.$$

The first term $\frac{\theta^A}{6\,k^2}$ represents the inconvenience costs associated with a perfect match $(a = u)$. The second term $\frac{\theta^A}{k^2}(a-u)^2$ captures the cost of a mismatch of quality d, with d as the distance metric $d \coloneqq |a-u|$.

Substituting d for $|a-u|$ yields the expected inter-category inconvenience costs $ic(d,k)$

$$E^P[ic(d)] = \frac{\theta^A}{6\,k^2} + \frac{\theta^A}{k^2}(d)^2.$$

From a mathematical perspective, the first term $\frac{\theta^A}{6\,k^2}$ is an additive constant, and $\frac{\theta^A}{k^2}$ is a multiplicative constant for every possible combination of categories a and u. Differences between inconvenience costs referring to single matches are then subject to a variation of d.

For a (complete) one-to-one matching that involves M couples, total inconvenience costs $TIC(d_i)$ are given by

$$E^P[TIC(d_i)] = M\frac{\theta^A}{6\,k^2} + \frac{\theta^A}{k^2}\sum_{i=1}^{M}(d_i)^2.$$

As noted above, the matching is efficient if total inconvenience costs are minimized. For a given a number of categories k, this is equivalent to minimizing the sum of squared distances $\sum_{i=1}^{M}(d_i)^2$. As $\sum_{i=1}^{M}(d_i)^2$ is determined by the assignment of category numbers a and u, the problem of efficient matching reduces to an allocation or coordination problem.

□ **Matching mechanism.** In this subsection, we present a matching mechanism based on Shen and Wong's (1983) work on histogram matching. Given a set of M users and M advertisers, our matching mechanism minimizes the sum of squared distances $D = \sum_{i=1}^{M}(d_i)^2$ between the two sets of agents. The resulting assignment is efficient in the sense that there is no other assignment or allocation generating lower total inconvenience costs.

The matching mechanism works as follows: After the subscription of M users and M advertisers, the platform operator observes the category number a of each advertiser and u of each user. The sets of observed category numbers are given by $\boldsymbol{a}$ and $\boldsymbol{u}$. As proven below, it takes two steps to produce an efficient assignment.

1. Sorting category numbers a_r and u_r in $\boldsymbol{a}$ and $\boldsymbol{u}$ such that the elements of $\boldsymbol{a} = \{a_1, a_2, \ldots, a_M\}$ and $\boldsymbol{u} = \{u_1, u_2, \ldots, u_M\}$ are both ordered in non-descending order $a_1 \leq a_2 \leq \cdots \leq a_{M-1} \leq a_M$ and $u_1 \leq u_2 \leq \cdots \leq$

$u_{M-1} \leq u_M$.[5] The subscript r denotes the r^{th} element in the ordered sets $\boldsymbol{a}$ and $\boldsymbol{u}$. In the following, we refer to r as the rank $r = \{1,2,\ldots,M\}$.

2. Matching category numbers element-wise. The resulting assignment of couples is the identity $\{(a_1,u_1),(a_2,u_2),\ldots,(a_M,u_M)\}$.

To prove the above claim, we first analyze a matching between two advertisements and two users ($|\boldsymbol{a}|,|\boldsymbol{u}| = M = 2$) and then generalize the findings for a larger matching.[6] Suppose set $\boldsymbol{a} = \{a_1, a_2\}$ is sorted, given that $a_1 \leq a_2$, and set $\boldsymbol{u}$ can either be sorted $\boldsymbol{u}^{\boldsymbol{sorted}} = \{u_1, u_2\}$ or not $\boldsymbol{u}^{\boldsymbol{not\ sorted}} = \{u_2, u_1\}$, given that $u_1 \leq u_2$. An element-wise matching leads to two possible assignments: $\{(a_1,u_1),(a_2,u_2)\}$ and $\{(a_1,u_2),(a_2,u_1)\}$. To check whether the sum of squared distances D is lower in the *sorted case* than in the *not sorted* case, we write

$$
\begin{aligned}
& D^{sorted} \leq D^{not\ sorted} \\
& <=> (a_1 - u_1)^2 + (a_2 - u_2)^2 \leq (a_1 - u_2)^2 + (a_2 - u_1)^2 \qquad (3) \\
& <=> (a_1 - a_2)\,(u_1 - u_2) \geq 0.
\end{aligned}
$$

The last inequality in Equation 3 is true if $u_1 \leq u_2 \wedge a_1 \leq a_2$ or $u_1 \geq u_2 \wedge a_1 \geq a_2$ is fulfilled. If $\boldsymbol{a}$ and $\boldsymbol{u}$ are sorted (in the same way), the element-wise matching with sorted category numbers always produces a distance (D^{sorted}) that is at least as small as the distance ($D^{not\ sorted}$) that results from an element-wise matching when

[5] Clearly, category numbers can also be sorted in a non-ascending order as long as both sets are ordered in the same way.

[6] The structure of the efficiency proof is equivalent to Melter et al. (1986), who deal with a linear matching and linear (absolute) distances.

$\boldsymbol{u}$ is not sorted. As a result: In a matching that includes two couples, elements in $\boldsymbol{u}^{not\ sorted}$ can always be exchanged $\left(\boldsymbol{u}^{not\ sorted} \to \boldsymbol{u}^{sorted}\right)$ without increasing the total inconvenience costs of the matching.

This result can be generalized for a larger matching that consists of $M > 2$ matches. Let us consider a set $\boldsymbol{a}$ that is sorted in a non-descending order and a set $\boldsymbol{u}$ with the sorting of category numbers u determined by an arbitrary permutation μ and $\boldsymbol{u} = \{u_{\mu 1}, u_{\mu 2}, \dots, u_{\mu M}\}$. The sum of squared distances for an element-wise matching is given by $D = \sum_{m=1}^{M}\left(a_m - u_{\mu\, m}\right)^2$, with m denoting the m^{th} match. If we now find an arbitrary subset of two couples with $u_{\mu\, m} \geq u_{\mu(m+1)}$, match partners can be exchanged and we get a new permutation with weakly lower total inconvenience costs. Starting from any initial sets $\boldsymbol{a}$ and $\boldsymbol{u}$ and continuing this way, the permutation μ will eventually become the identity where a_r is matched with u_r, with $r = m$ (see: Melter et al, 1986).[7]

TABLE 1

The following example illustrates the proof of efficiency: Suppose the number of subscribers on both sides of the market is $M = 6$ and the platform operator identifies $k = 3$ categories. Suppose the sorted set $\boldsymbol{a} = \{1,1,2,2,3,3\}$ and the permutation $\boldsymbol{u} = \{2,2,1,3,1,3\}$ represent the respective category numbers of each user and advertiser active on the platform. The gradual steps from the initial assignment to the efficient assignment are depicted in Table 1.

[7] In terms of bipartite graph theory, this resulting assignment would contain a *perfect bipartite matching* (see e.g., Easley and Kleinberg, 2010).

Given the initial assignment, look for the first appearance(s) of $u_{\mu\,m} \geq u_{\mu(m+1)}$. In the initial set $\boldsymbol{u}$ represented in column 1, this is true for the subsets involving matches $m = 2{,}3$ and matches $m = 4{,}5$. For both subsets, exchanges can be performed as indicated. Distance is reduced by $\Delta D = (1^2 + 1^2 + 1^2 + 2^2) - (0^2 + 0^2 + 1^2 + 0^2) = 6$. The second column shows the resulting assignment determined by a new permutation μ in $\boldsymbol{u}$. Exchanges in subsets including matches $m = 1{,}2$ and $m = 3{,}4$ are made and result in a change of the total distance: $\Delta D = (1^2 + 0^2 + 0^2 + 1^2) - (0^2 + 1^2 + 1^2 + 0^2) = 0$. In the third column, subset $m = 2{,}3$ is critical. Exchanging match partners again reduces the total distance by $\Delta D = (1^2 + 1^2) - (0^2 + 0^2) = 2$. Given this permutation, it is not possible to find an assignment of advertisements to users' eyeballs that is more efficient than this.

By assumption, advertisers are informed about the matching mechanism. As they do not ex ante know the exact allocation of users over categories, but only the distribution of users in the population, they can only calculate the probabilities $w(a, u)$ for a match with each specific user category u before subscribing to the platform service. In the next section, we derive these probabilities.

□ **Match probabilities.** As stated above, we assume that there are M advertisers and M users active on the platform. We further assume that the distribution of advertisements over the k categories is deterministic and the number of advertisements in each category is $\frac{1}{k}M$. This assumption is made for simplification, but also captures the planned and longer-term character of advertising campaigns, which makes the demand for intermediation predictable for the platform operator ex ante.

In contrast to advertisers, users with ex ante random preferences join the platform service. Thus, the user-specific category number u is a random variable. The probability of realizations of u is derived from the probability density distribution of preferences x^U in the population. Remember from above that user preferences x^U are uniformly distributed over a line segment $\Psi \coloneqq [0,1]$, so the probability density function and the cumulative distribution function are given by $f(x^U) = 1$ and $F(x^U) = x^U$respectively, as shown in Figure 3a) and Figure 3b).

FIGURE 3

As the platform operator can only identify the category u of a user, we transform the continuous distribution function of characteristics into a discrete distribution of categories: $f(x^U) = 1 \rightarrow f^U(u) = \frac{1}{k}$, as shown in Figure 3c). The corresponding cumulative distribution function $F^U(u) = \frac{u}{k}$ defines the probability of user i subscribing to a category number equal to or smaller than u (Figure 3d)).

To derive the match probabilities $w(a,u)$, which enter the advertiser's utility function as weights, we proceed in two steps. First, we derive a rank-based probability $w(r,u)$, i.e., the probability of an advertiser positioned on rank r in $\boldsymbol{a}$ being matched to a user of a specific category u. Second, we derive the probabilities $w(a,u)$ for a match between a specific advertiser of category a and a specific user of category u.

Rank-based match probabilities. Rank-based match probabilities $w(r,u)$ are given by[8]

$$w(r,u) = \sum_{h=r}^{M} \binom{M}{h} \left[\left(F^U(u)\right)^h \left(1 - F^U(u)\right)^{M-h} - \left(F^U(u^-)\right)^h \left(1 - F^U(u^-)\right)^{M-h} \right]. \qquad (4)$$

To see this, note that for matches that include a user of category u and an advertiser on a specific rank r, there are two kinds of conditions. First, the number of subscribers h of a category smaller than or equal to category number u is **at least as high** as the advertiser's rank r. If this condition were not fulfilled, it would not be possible to match an advertiser on a specific $rank$ with a user of exactly category u. The respective probability reads as

$$P(\, h \geq r \; subscribers \; in \; \boldsymbol{u} \; with \; a \; category \; number \; \boldsymbol{lower \; or \; equal} \; to \; u)$$

$$= \sum_{h=r}^{M} \binom{M}{h} \left(F^U(u)\right)^h \left(1 - F^U(u)\right)^{M-h}.$$

Each distinct probability of exactly h subscribers (in the sample) is given by the binomial probability $\binom{M}{h} * \left(F^U(u)\right)^h \left(1 - F^U(u)\right)^{M-h}$, with $\binom{M}{h}$ as the number of permutations that can possibly be rearranged to the same sorted set of category numbers $\boldsymbol{u}$. The above probability for at least $h \geq r$ subscribers of category u or lower is determined by the sum of all distinct probabilities that correspond to cases of exactly $h \wedge h + 1 \wedge ... \wedge M$ users of these categories in the sample.

[8] See Arnold, Balakrishnan, and Nagaraja (1992) for detailed information on order statistics.

The second condition for a match between a user of a specific category number u and an advertiser assigned to rank r is that the number of subscribers h of a category strictly lower than category number u is **strictly lower** than the advertiser's rank r. This condition rules out the case of a category number lower than u (denoted as u^-) being assigned to the specific rank r. We therefore subtract the corresponding probabilities so that an advertiser j on r is exactly matched to a user i of category number u:

$$P(h \geq r \; subscribers \; in \; \boldsymbol{u} \; with \; a \; category \; number \; \boldsymbol{strictly \; lower} \; than \; u)$$

$$= \sum_{h=r}^{M} \binom{M}{h} \left(F^U(u^-)\right)^h \left(1 - F^U(u^-)\right)^{M-h}.$$

Equation 4 merges both conditions.

Category-based match probabilities. To get from the rank-based probabilities to the category-based match probabilities, we calculate the simple average of rank-based probabilities that fall into the same category. With several advertisers from a certain category a, each advertiser in a has the same probability to be assigned to a user in category u. Given our assumption of deterministic entry, the number of advertisers in all categories is the same and equal to $\frac{M}{k}$. Therefore, category-based match probabilities simplify to

$$w(a,u) = \frac{k}{M} \sum_{r=M\frac{(a-1)}{k}+1}^{M\frac{a}{k}} w(r,u).$$

Note that ranks r that correspond to a specific category a start from $r = \mathrm{M}\,\frac{(a-1)}{k} + 1$ and end at $r = M\,\frac{a}{k}$. The sum of these rank-based probabilities divided by the number of advertisers in each category $\frac{M}{k}$ yields the simple average of rank-based probabilities in each category, i.e., the category-based probability.

Figure 4a) shows probabilities for $k = 3$ categories and $M = 3$ subscribers, whereas Figure 4b) shows probabilities for $k = 3$ categories and $M = 90$ subscribers. It can be seen that the probabilities for perfect matches $w(a,u|a = u)$ increase with M, whereas the probabilities for imperfect matches $w(a,u|a \neq u)$ decrease with M.

FIGURE 4

Match probabilities may be used to derive a micro-foundation for cross-side externalities. Analogous to Rochet and Tirole (2006), let the interaction parameter β^A denote the average benefit a representative agent (advertiser) receives from an average interaction. Cross-side externalities in Rochet and Tirole (2006) derive from β^A multiplied by M possible interactions. In our model, there is only one interaction or match that is mediated by the platform operator. Thus, the cross-side effect $\beta^A(M)$ is a function of possible matches. Cross-side externalities read as

$$\beta^A = V^A - \frac{\theta^A}{k^3}\sum_{a=1}^{k}\sum_{u=1}^{k}\left(w(a,u)\int_{\frac{a-1}{k}}^{\frac{a}{k}}\left(\int_{\frac{u-1}{k}}^{\frac{u}{k}}(x_u^U - x_a^A)^2 f(x_u^U)\, d\, x_u^U\right) f(x_a^A)\, d\, x_a^A\right)$$

$$<=> \; V^A - \frac{\theta^A}{6\,k^2} - \frac{\theta^A}{k^2}\sum_{a=1}^{k}\sum_{u=1}^{k}\frac{w(a,u)\,(a-u)^2}{k}.$$

Given the reservation value for an ideal match V^A, the average benefit from interaction is reduced due to the dispersion of preferences and characteristics within categories $\frac{\theta^A}{6\,k^2}$ and the average mismatching between categories $\frac{\theta^A}{k^2}\sum_{a=1}^{k}\sum_{u=1}^{k}\frac{w(a,u)\,(a-u)^2}{k}$. As illustrated in Figure 5, cross-side externalities may be negative for some parameterizations. The strength of cross-side externalities increases with the number of potential match partners due to the increasing probability of being involved in a high-quality match. Starting from a small scale, interaction benefits increase less than proportionally with the number of match partners and eventually converge to $\frac{\theta^A}{6\,k^2}$, i.e., the disutility caused by inaccurate identification when the platform operator matches perfect couples only. We later refer to this situation as the benchmark of deterministic subscription to the platform service.

FIGURE 5

□ **Expected inconvenience costs.** Expected inconvenience costs $E^A\left[ic\left(a, x_j^A, \theta^A\right)\right]$ of an individual advertiser positioned on x_j^A are now fully specified as

$$E^A\left[ic\left(a, x_j^A, \theta^A\right)\right] = \theta^A \sum_{u=1}^{k} \left(w(a,u) \int_{\frac{u-1}{k}}^{\frac{u}{k}} \left(x_j^A - x_u^U\right)^2 f(x_u^U) \, \mathrm{d}\, x_u^U \right).$$

They derive from the weighted sum of the expected cost of a match with a user of a certain category.

FIGURE 6

Figure 6 plots inconvenience costs for each brand positioning x_j^A for a parameterization of $M = 6$ and $k = 3$ categories in $6a$) and $M = 6$ and $k = 6$ categories in $6b$). As depicted, the expected inconvenience cost curve $E^A\left[ic\left(a, x_j^A, \theta^A\right)\right]$ has three important properties: A convex shape within categories, jump discontinuities between categories, and symmetry with respect to the center.

The convex shape of $E^A\left[ic\left(a, x_j^A, \theta^A\right)\right]$ within categories a is due to squared distance costs. This means that the maximum of inconvenience costs in each category is located either on the left boundary $\left(\frac{a-1}{k}\right)$ or on the right boundary $\left(\frac{a}{k}\right)$ of each category interval ψ_a^A. Jump discontinuities are due to the categorization and sorting performed by the platform operator. The jump discontinuities decrease with the number of subscribers and totally disappear for $M \to \infty$. The total maximum (maxima) of inconvenience costs is (are) either located on the lateral boundary (Figure 6b)) of the centric interval(s) or on the end points of $\Psi \coloneqq [0,1]$ (Figure 6a). We observe two effects that describe the phenomenon. First, there is a sorting effect that increases the likelihood of perfect matches the closer advertisers are to the end points. These advertisers have a higher probability to obtain a high-quality match compared with advertisers assigned to center categories. Secondly, we observe a distance effect. As inconvenience costs increase more than proportionally with distance, advertisers with a brand positioning near the center incur lower inconvenience costs. The third property of the expected inconvenience cost curve is a symmetric shape. If we draw a vertical line at $x = \frac{1}{2}$, the two sides left and right of the center are identical.

Figure 6 shows the impact of more accurate identification on the level of inconvenience costs. Increasing k from three to six categories decreases inconvenience costs significantly. This is illustrated by the vertical difference between the value of an ideal match V^A and the inconvenience cost curve $E^A\left[ic\left(a, x_j^A, \theta^A\right)\right]$.

5. Incentives to invest

■ In this section, we derive the profit-maximizing level of identification. We first set up the profit function for what we denote as deterministic subscription of users. This constitutes a benchmark abstracting from mismatches between categories due to subscription of random users. Then we describe the profit maximization problem for the case in which random users join the platform service and the discrete distribution of users over categories results from a stochastic process. The platform must decide how to coordinate imperfect couples. Here, the matching mechanism and the impact of agent-specific information on match efficiency come into play. Distinguishing between the two cases allows us to isolate two opposing effects that determine the incentives to invest in technologies generating agent-specific information:

1. Scalability of fixed investments into the identification technology. This effect works in favor of platforms with large customer bases.
2. Match efficiency. Efficiency gains are generated in the allocation process. More precise user-specific information leads to a more efficient reallocation of match partners. This effect is more significant for platforms with small customer bases.

To derive investment incentives, we set up the profit function

$$\pi = (p^A(k) - c)\, M - k\, C.$$

$p^A(k)$ denotes the price charged from advertisers for a single match. Marginal costs are denoted by c. In case of full market coverage, M denotes the demand for intermediation and $(p^A(k)\ - c)\, M$ the revenue. The cost of identification is given by $k\, C$, with C as the constant costs of identifying an additional category.

□ **Benchmark model.** As mentioned above, we assume full market coverage on the user side: No matter how inconvenient, advertisements will never prevent users from using the platform service. Thus, the market is covered on the user side by assumption. For advertisers, we assume full market coverage as well, which implies a weakly positive expected utility on their part. Demand of users and advertisers for the platform service is therefore given by M. We assume a deterministic and uniform distribution of advertisers and users over categories in the benchmark model. Thus, demand for each category is given by $q_a^A = (F^A\left(\frac{a}{k}\right) - F^A\left(\frac{a-1}{k}\right))M$ and $q_u^U = (F^U\left(\frac{u}{k}\right) - F^U\left(\frac{u-1}{k}\right))M$.

FIGURE 7

The inconvenience cost curves for the benchmark case are depicted by the solid lines in Figure 7. The inconvenience cost curve is symmetric in every category, and jump discontinuities do not exist due to the matching of perfect couples only. The distance effect determines the convex shape of the inconvenience cost curve only within a category but not between categories.

Profit-maximization. Under the assumption of full market coverage, profit-maximizing prices are endogenous. The highest price a platform operator can charge for the subscription to each category is the one that leaves the producer of the pivotal advertisement ${x_j^A}^{pivotal}$ with a net utility of 0. The pivotal brand positioning ${x_j^A}^{pivotal}$ is the one that fits least with the preferences of expected match partner i under the given matching mechanism. This specific advertiser has the highest inconvenience costs in a category. Due to the convex shape of the inconvenience cost curve, pivotal advertisers ${x_j^A}^{pivotal}$ are either positioned on the left $\left(\frac{a-1}{k}\right)$ or on the right boundary $\left(\frac{a}{k}\right)$ of ψ_a^A. As the inconvenience cost curve is fully symmetric in each category interval, maximal inconvenience costs are the same for positions $\left(\frac{a-1}{k}\right)$ and $\left(\frac{a}{k}\right)$ in a certain category a and between categories (See Figure 7). The platform operator charges a uniform price and does not price discriminate between categories.

The net utility of a pivotal advertiser is

$$E^A\left(v\left(\frac{a-1}{k}\right)\right) = E^A\left(v\left(\frac{a}{k}\right)\right)$$

$$= V^A - p^A - \theta^A \left(\int_{\frac{u-1}{k}}^{\frac{u}{k}} \left(\left(\frac{a-1}{k}\right) - x_u^U\right)^2 f(x_u^U)\, \mathrm{d}\, x_u^U\right). \quad (5)$$

Equation 5 is a special case of Equation 1. The probability $w(a,u)$ for perfect matches is 1. Equation 5 simplifies to $V^A - p^A - \theta^A \frac{1}{3\,k^2}$. Setting the utility of the pivotal advertiser equal to 0 and solving for p^A yields the profit-maximizing price

$$p^{A^*}(k) = V^A - \theta^A \frac{1}{3\,k^2}.$$

Optimal level of identification. Given prices p^{A^*}, the reduced profit function reads

$$\pi(k) = \left(V^A - \theta^A \frac{1}{3\,k^2} - c\right) M - k\,C.$$

For an exogenous market size, the platform operator maximizes profits by minimizing the sum of inconvenience costs and identification costs. The first-order condition reads as

$$\frac{d\,\pi(k)}{d\,k} = \mathrm{M} \frac{2}{3} \frac{\theta^A}{k^3} - C = 0. \quad (6)$$

Equation 6 shows that the optimal level of agent-specific information is reached when the costs of introducing an additional category C equal the revenue gained from introducing an additional category $M\left(\frac{2}{3}\frac{\theta^A}{k^3}\right)$, with $\frac{2}{3}\frac{\theta^A}{k^3}$ as the decrease in inconvenience costs of mismatching. Solving the first-order condition for k yields

$$k^{opt} = \left(\frac{2}{3}\frac{M\,\theta^A}{C}\right)^{\frac{1}{3}}. \tag{7}$$

It can be seen from Equation 7 that k^{opt} increases with the sensitivity for imperfect match partners θ^A $\left(\frac{\delta\, k^{opt}}{\delta\, \theta^A} > 0\right)$ and with market size $\left(\frac{\delta\, k^{opt}}{\delta\, M} > 0\right)$. The more users and advertisers are matched, the higher the marginal revenue generated by more accurate identification. The level of identification decreases in the costs C for identifying an additional category $\left(\frac{\delta\, k^{opt}}{\delta\, C} < 0\right)$. The exponent $\frac{1}{3}$ derives from the unit interval $\Psi \coloneqq [0,1]$. An additional category k decreases the absolute length of a representative subinterval ψ_u^U and ψ_a^A less than proportionally, whereas costs increase by the constant amount C.

□ **Mark-ups and incentives to invest.** We now derive the impact of identifying an additional category ($\Delta_k = 1$) on profitability as the relative strength of incentives to invest into identification for the benchmark case given by $\frac{\Delta_k \pi^{det}}{k} = \frac{\pi^{det}(k+1) - \pi^{det}(k)}{k}$. The expression simplifies to

$$\Delta_k P^{det} - \Delta_k AC^{det} = \frac{(2k+1)\,\theta^A}{3k^2(k+1)^2} - \frac{C}{M} \quad (8)$$

and provides the additional absolute mark-up $\Delta_k P^{det} - \Delta_k AC^{det}$ gained from increasing the accuracy of identification by one additional category Δ_k. The uniform price increases by $\Delta_k P^{det} = \frac{2}{3}\frac{\theta^A}{k^3}$, and the costs per match increase by $\Delta_k AC^{det} = \frac{C}{M}$. The impact on profits, and thus on incentives to invest, is linear in the (additional) mark-up because market size is exogenous and the market is assumed to be fully covered. Thus, there is no quantity effect resulting from $\Delta_k P^{det}$. Equation 8 therefore measures the profitability of increasing the accuracy level of identification by one category due to economies of scale and the reduction of dispersion within a category.

FIGURE 8

Figure 8 shows the change in prices $\Delta_k P$ and the change in costs per match $\Delta_k AC$ due to a marginal increase in the accuracy of identification (ordinate) as a function of market size (abscissa). The solid line represents the increase in price $\Delta_k P$. It is constant because the probability for perfect matches $w(a, u|a = u)$ is equal to 1 given deterministic subscription. The dotted line represents the cost per match $\Delta_k AC$ of identifying an additional category. The intersection between the two curves shows the minimal scale required for making the investment profitable.

□ **Profit with random subscribers.** In this section, we analyze profit maximization in a model where random users subscribe to the platform. In this setting, our matching mechanism distributes advertisements to eyeballs. We focus on the impact of identification on match quality and show that efficiency gains are greater for platforms with a smaller customer base.

Profit function. We again assume full market coverage and price differentiation between categories. The platform operator charges prices p_a^A for placing an advertisement in category a. This price fully extracts the net utility of the advertiser who places the pivotal advertisement ${x_{j,a}^A}^{pivotal}$. The pivotal advertisement ${x_{j,a}^A}^{pivotal}$ is located on the edge of each subinterval ψ_a^A. To derive the profit-maximizing price ${p_a^A}^*$ for an advertisement of category a and the resulting revenue R^A, we have to distinguish between the revenue for even and odd numbers of categories k.

Revenue $R^A(k)$ for an <u>even</u> number of categories k on the advertiser side reads as

$$R^A(k) = \sum_{a=1}^{k} p_a^A \, q_a^A = 2 \sum_{a=1}^{\frac{k}{2}} p_a^A \frac{M}{k} \, .$$

Revenue $R^A(k)$ is given by the sum of revenues earned from advertisements of each category $p_a^A \, q_a^A$. The demand for placing an advertisement in category a is $\frac{M}{k}$. Due to the symmetry of the inconvenience cost curve, revenues are $2\sum_{a=1}^{\frac{k}{2}} p_a^A \frac{M}{k}$. By summing up revenues in categories left of the center and multiplying the sum by 2, we can disregard the switching of pivotal positions ${x_{j,a}^{A}}^{pivotal}$ from the left to the right boundary when category numbers are greater than $a = \frac{k}{2}$ (see Figure 6).

Setting the net utility of the pivotal advertiser on ${x_{j,a}^{A}}^{pivotal} = \frac{a-1}{k}$ in each category a equal to 0 and solving for the price, we get $p_a^A = V^A - \theta^A \sum_{u=1}^{k} \left(w(a,u) \int_{\frac{u-1}{k}}^{\frac{u}{k}} \left(\frac{a-1}{k} - x_u^U \right)^2 f(x_u^U) \, \mathrm{d} x_u^U \right)$.

Substituting prices p_a^A into the revenue function eventually yields

$$R^A = 2 \sum_{a=1}^{\frac{k}{2}} \left(V^A - \theta^A \sum_{u=1}^{k} \left(w(a,u) \int_{\frac{u-1}{k}}^{\frac{u}{k}} \left(\frac{a-1}{k} - x_u^U \right)^2 f(x_u^U) \, \mathrm{d} x_u^U \right) \right) \frac{M}{k}.$$

Revenue $R^A(k)$ for an <u>odd</u> number of categories k reads as

$$R^A(k) = 2 \sum_{a=1}^{\frac{k-1}{2}} \left(p_a^A \frac{M}{k} \right) + p_{\frac{k+1}{2}}^A \frac{M}{k}$$

$$= 2 \sum_{a=1}^{\frac{k-1}{2}} \left(V^A - \theta^A \sum_{u=1}^{k} \left(w(u,a) \int_{\frac{u-1}{k}}^{\frac{u}{k}} \left(\frac{a-1}{k} - x_u^U \right)^2 f(x_u^U) \, \mathrm{d} x_u^U \right) \right) \frac{M}{k}$$

$$+ \left(V^A - \theta^A \sum_{u=1}^{k} \left(w\left(\frac{k+1}{2}, u \right) \int_{\frac{u-1}{k}}^{\frac{u}{k}} \left(\frac{k+1}{2\,k} - x_u^U \right)^2 f(x_u^U) \, \mathrm{d} x_u^U \right) \right) \frac{M}{k}.$$

The first term $2\sum_{a=1}^{\frac{k-1}{2}}\left(p_a^A \frac{M}{k}\right)$ sums up revenues in each category strictly left of $x = \frac{1}{2}$, excluding the center category, and doubles them. The second term $p_{\frac{k+1}{2}}^A \frac{M}{k}$ represents the revenue earned in the center category $\frac{k+1}{2}$. Prices p_a^A for both even and odd numbers of categories k are illustrated in Figure 8.[9]

FIGURE 9

Platform profit $\pi(k)$ is given by

$$\pi(k) = R^A(k) - k\,C.$$

□ **Mark-ups and efficiency gains from identification.** We are now able to isolate the match-efficiency effect. It is depicted in the Figure below.

FIGURE 10

Figure 10 shows the change in mark-ups (ordinate) due to the identification of one additional category as a function of market size (abscissa). The solid line depicts changes in the price $\Delta_k P^{det}$ for the benchmark of deterministic platform subscription. The dashed line depicts changes in the average price $\Delta_k P^{random}$ for the case in which subscribers with random preferences join the platform. The shaded area between these two curves shows the impact of acquiring additional information on more efficient allocation of match partners and the corresponding increase in mark-ups. Analogous to Figure 8, the dotted line shows the costs per match of identifying an additional category.

[9] Figure 9a) shows that specific advertisers might have incentives to misrepresent their brand positioning ($\frac{1}{3} - \varepsilon \to \frac{1}{3} + \varepsilon$) in order to pay lower prices ($p_1^A \to p_2^A$) for the platform service. This may be of practical relevance as there are many consultancies optimizing keyword strategies etc.

We see from Figure 10 that efficiency gains correlate positively with the probability of imperfect matches and negatively with market size. This means that additional user-specific information is more beneficial for the coordination process between advertisers and users in smaller markets. Stochastic platform usage coincides with a relatively high probability for imperfect matches in small markets and a rather low probability for such matches in large markets. Additional agent-specific information leads to a more efficient reallocation of pairs that were seen as optimally matched before, and matches of the lowest quality are avoided.

Firms are willing to make the additional investment on smaller scales if there is uncertainty about the distribution of users over categories. To see this from Figure 10, compare the minimal market size required to make the additional investment profitable (break-even scale) in the random model (i.e., the interception of $\Delta_k AC$ (dotted line) and $\Delta_k P^{random}$ (dashed line)) with the break-even scale in the deterministic model (i.e., the interception of the $\Delta_k AC$ (dotted line) and $\Delta_k P^{det}$ (solid line)).

6. Discussion

■ Our article analyzes the incentives to invest in acquiring information on agent-specific characteristics and preferences for a monopolistic matchmaker in a two-sided market. In a model where agents have assortative preferences for match partners, we derive the following results: There is a profit-maximizing accuracy level of agent-specific information determined by a trade-off between higher prices due to higher expected match quality on the one hand and incurring fixed costs for identification technology on the other. We also show that the impact of more accurate identification on match efficiency is higher for platforms with smaller customer bases. This effect is interesting to isolate because it works in the opposite direction of economies of scale.

We define match efficiency as the probability of being involved in a high-quality match. This definition allows us to derive a micro-foundation for the strength of cross-side externalities, and it provides a link to the seminal articles in the literature on two-sided markets. Our main contribution is that we derive the expected willingness to pay for match quality in a closed form, thus doing away with the need to simulate matching outcomes.

Deriving these results in a closed form requires some restrictive assumptions. In our model, agents are uniformly distributed over a one-dimensional spectrum of characteristics, incur quadratic inconvenience costs of mismatching, and their demand fully covers the market. These assumptions, however, open up an avenue for further research. One especially promising question may be whether platform operators – both in a monopolistic and in a competitive environment – should provide their matching services only to a particular segment of potential customers.

We also do not focus on advertisers' incentives to plan their advertising campaigns strategically and place advertisement that deviate from their brand positioning in an effort to save costs. Analyzing this strategic aspect of advertisers' behavior is of high practical relevance and appears to be a particularly interesting task for future research.

References

Armstrong, M. "Competition in Two-Sided Markets." *RAND Journal of Economics*, Vol. 37 (2006), pp. 668-691.

Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N., eds. *A First Course in Order Statistics*. New York: Wiley, 1992.

Damiano, E. and Li, H. "Price Discrimination and Efficient Matching." *Economic Theory*, Vol. 30 (2007), pp. 243-263.

Demange, G., Gale, D., and Sotomayor, M. (1986) "Multi-Item Auctions." *The Journal of Political Economy*, Vol. 94 (1986), pp. 863-872.

Easley, D. and Kleinberg, J., eds. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge: Cambridge University Press, 2010.

Evans, D.S. and Schmalensee, R. "The Industrial Organization of Markets with Two-Sided Platforms." *Competition Policy International*, Vol. 3 (2007), pp. 151-179.

Gale, D. and Shapley, L. "College Admissions and the Stability of Marriage." *American Mathematical Monthly*, Vol. 69 (1962), pp. 9-15.

Hitsch, G.J., Hortaçsu, A., and Ariely, D. "Matching and Sorting in Online Dating." *The American Economic Review*, Vol. 100 (2010), pp. 130-163.

Hotelling, H. "Stability in Competition." *Economic Journal*, Vol. 39 (1929), pp. 41-57.

Kalmijn, M. "Intermarriage and Homogamy: Causes, Patterns, Trends." *Annual Review of Sociology*, Vol. 24 (1998), pp. 395–421.

Kuhn, H.W. "The Hungarian Method for the Assignment Problem." *Naval Research Logistics Quarterly*, Vol. 2 (1955), pp. 83–87.

Lee, R. and Schwarz, M. "Interviewing in Two-Sided Matching Markets." NBER Working Paper no. 14922, Harvard University, National Bureau of Economic Research, 2012.

Munkres, J.R. "Algorithms for the Assignment and Transportation Problems." *Journal of the Society of Industrial and Applied Mathematics*, Vol. 5 (1957), pp. 32–38.

Parker, G.G. and Van Alstyne, M.W. "Two-Sided Network Effects: A Theory of Information Product Design." *Management Science*, Vol. 51 (2005), pp. 1494-1504.

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

Rochet, J.-C. and Tirole, J. "Two-Sided Markets: A Progress Report." *RAND Journal of Economics*, Vol. 37 (2006), pp. 645–667.

Shen, H.C. and Wong, A.K.C. "Generalized Texture Representation and Metric." *Computer Visions, Graphics and Image Processing*, Vol. 23 (1983), pp. 187-206.

Shimer, R. and Smith, L. "Assortative Matching and Search." *Econometrica*, Vol. 68 (2000), pp. 343-369.

Melter, R., Peleg, S., Rosenfeld, R., and Werman, M. "Bipartite Graph Matching for Points on a Line or a Circle." *Journal of Algorithms*, Vol. 7 (1986), pp. 277-284.

Weyl, E.G. "The Price Theory of Multi-Sided Platforms." *The American Economic Review*, Vol. 100 (2010), pp. 1642-72.

Figures

FIGURE 1

THREE CATEGORIES OF USERS AND ADVERTISERS

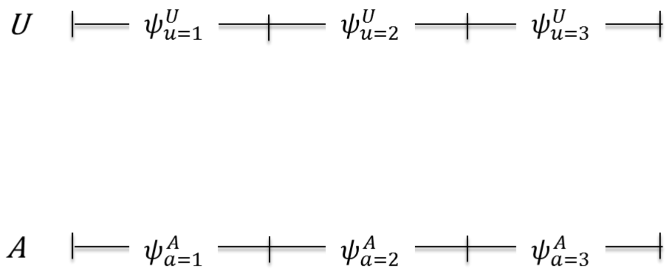


FIGURE 2

ILLUSTRATION OF MATCH PROBABILITIES

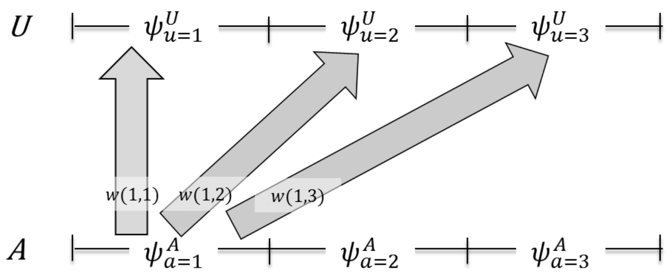


FIGURE 3

TRANSFORMATION OF THE CONTINUOUS DISTRIBUTION OF PREFERENCES INTO DISCRETE DISTRIBUTION OF CATEGORY NUMBERS

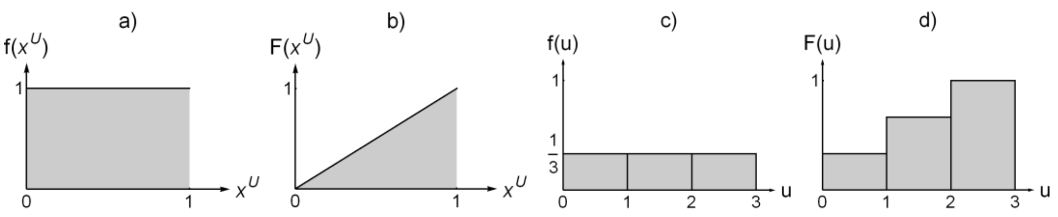


FIGURE 4

DISTRIBUTIONS OF MATCH PROBABILITIES

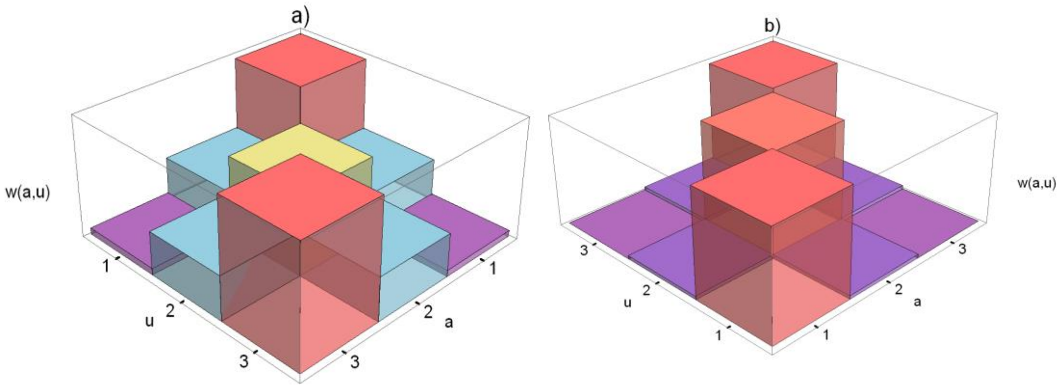


FIGURE 5

STRENGTH OF CROSS-SIDE EXTERNALITIES AS A FUNCTION OF MARKET SIZE

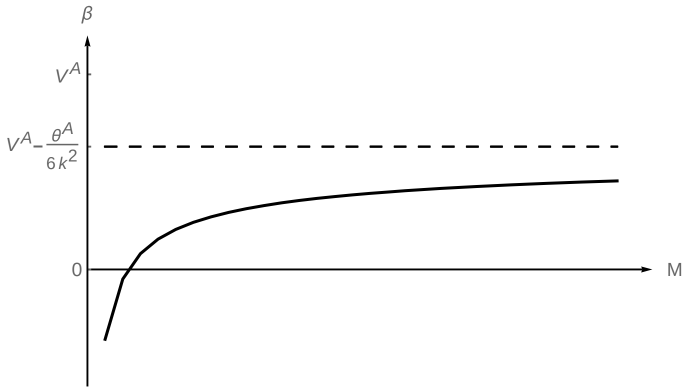


FIGURE 6

PROPERTIES OF THE EXPECTED INCONVENIENCE COST CURVE

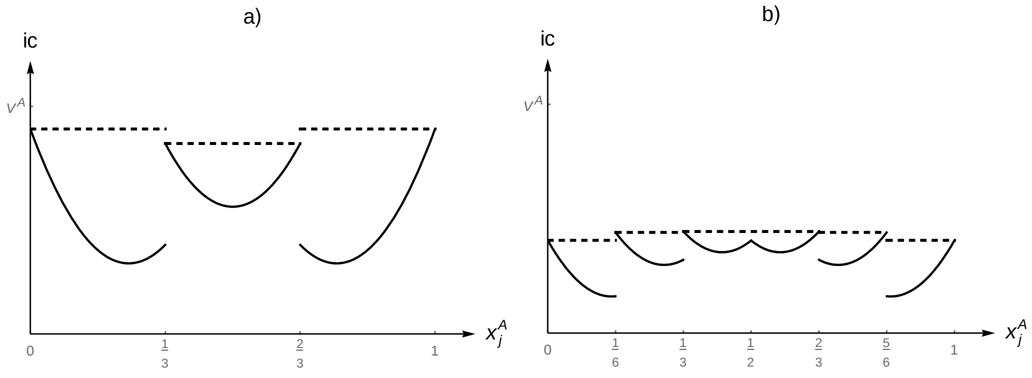


FIGURE 7

EXPECTED INCONVENIENCE COSTS IN THE BENCHMARK CASE

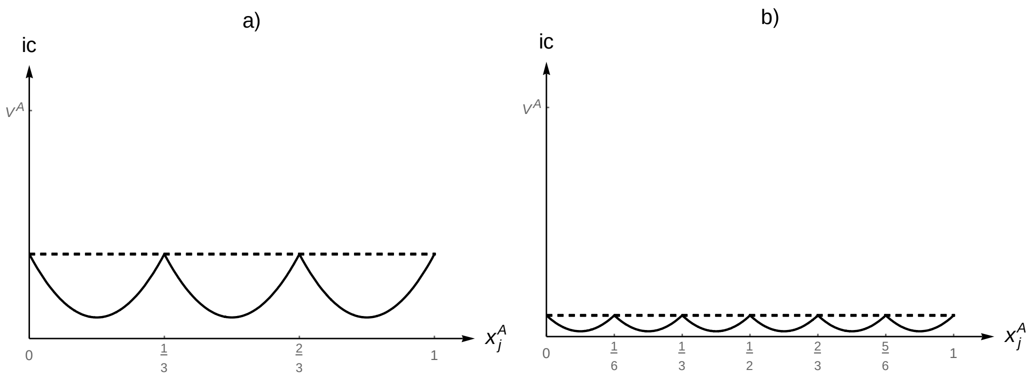


FIGURE 8

PROFITABILITY OF INVESTMENTS IN THE BENCHMARK CASE

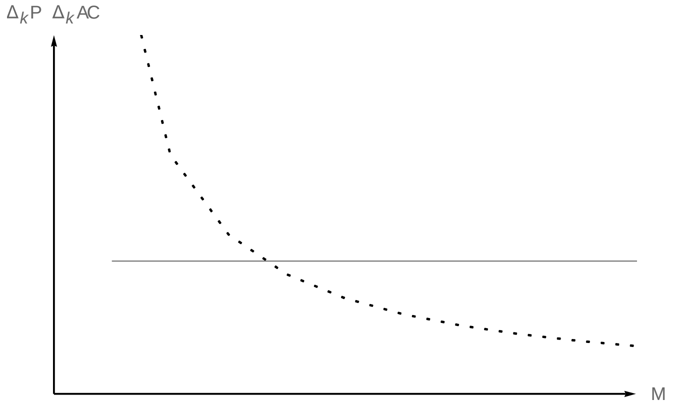


FIGURE 9

PRICING STRUCTURE ON ADVERTISER SIDE

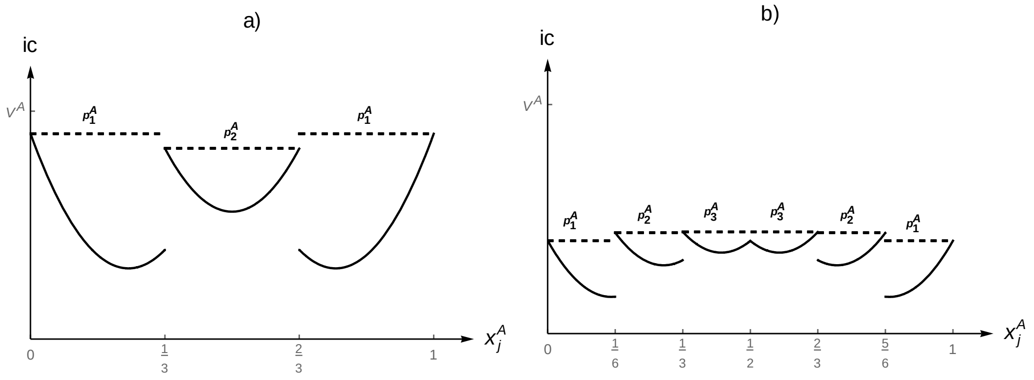
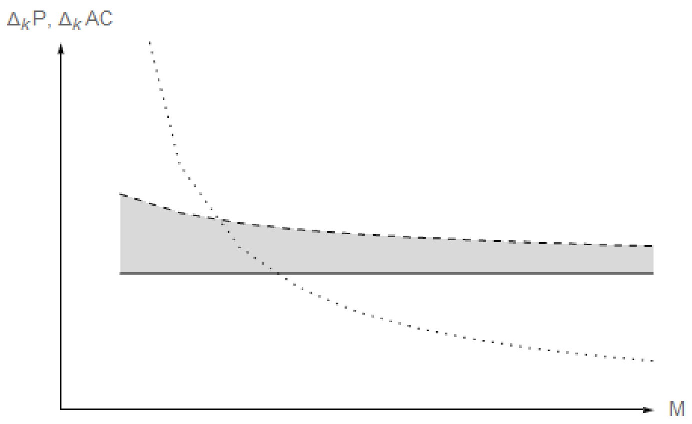


FIGURE 10

PROFITABILITY OF INVESTMENTS



Tables

TABLE 1

SORTING AND EFFICIENCY OF MATCHING

$\boldsymbol{d_m =}$	$\boldsymbol{(a_m - u_{\mu m})^2}$	$\boldsymbol{(a_m - u_{\mu m})^2}$	$\boldsymbol{(a_m - u_{\mu m})^2}$	$\boldsymbol{(a_m - u_{\mu m})^2}$
$d_1 =$	$(1 - 2)^2$	$(1 - 2)^2$	$(1 - 1)^2$	$(1 - 1)^2$
$d_2 =$	$(1 - 2)^2$	$(1 - 1)^2$	$(1 - 2)^2$	$(1 - 1)^2$
$d_3 =$	$(2 - 1)^2$	$(2 - 2)^2$	$(2 - 1)^2$	$(2 - 2)^2$
$d_4 =$	$(2 \quad 3)^2$	$(2 - 1)^2$	$(2 - 2)^2$	$(2 - 2)^2$
$d_5 =$	$(3 - 1)^2$	$(3 - 3)^2$	$(3 - 3)^2$	$(3 - 3)^2$
$d_6 =$	$(3 - 3)^2$	$(3 - 3)^2$	$(3 - 3)^2$	$(3 - 3)^2$