To match or not to match: Economics of cookie matching in online advertising

Arpita Ghosh, Cornell University
Mohammad Mahdian, Google Inc.
R. Preston McAfee, Google Inc.
Sergei Vassilvitskii, Google Inc.

Modern online advertising increasingly relies on the availability of user tracking technology called *cookie-matching* to increase efficiency in ad allocation. Web publishers today use this technology to share information about the websites a user has visited, making it possible to target advertisements to users based on their prior history. This begs the question: do publishers (who are competitors for advertising money) always have the incentive to share online information? Intuitive arguments as well as anecdotal evidence suggest that sometimes a premium publisher might suffer from information sharing through an effect called *information leakage*: by sharing user information with the advertiser, the advertiser will be able to target the same user elsewhere on cheaper publishers, leading to a dilution of the value of the supply on the premium publishers.

The goal of this paper is to explore this aspect of online information sharing. We show that when advertisers are homogeneous, in the sense that their relative valuations of users are consistent, publishers always agree about the benefits of cookie-matching *in equilibrium*: either all publishers' revenues benefit, or all suffer, from cookie-matching. We also show using a simple model that when advertisers are not homogeneous, the information leakage indeed can occur, with cookie-matching helping one publisher's revenues while harming the other.

Categories and Subject Descriptors: J.4 [**Social and Behaviorial Sciences**]: Economics

General Terms: Economics, Theory, Algorithms

Additional Key Words and Phrases: Ad Auctions, Cookie-matching, Information in Auctions

1. INTRODUCTION

When will competitors share online information? We consider this question in the context of Internet cookies, which are small files placed on a user's computer that permit a website to record information about a previous visit. Cookies can be used to provide a better user experience, e.g., by allowing a user to stay logged in on a website or by remembering user preferences, and can also be used to target advertising.

Many websites now share cookie information with each other [Perlich and Dalessandro 2013]. For instance, a user might notice that after she searches for flights to Hawaii on Orbitz.com, ads relevant to Hawaii continue to follow her across the web, showing up, say, when she visits the New York Times website. While such cookie sharing

---

Author's email addresses: Arpita Ghosh, arpitaghosh@cornell.edu; Mohammad Mahdian, mahdian@alum.mit.edu; R. Preston McAfee, preston@mcafee.cc; Sergei Vassilvitskii, sergei@cs.stanford.edu.

EC'12, June 4–8, 2012, Valencia, Spain.

amongst websites creates some obvious conveniences for the user, it also permits more *targeted* advertising by creating a more detailed picture of the customer, leading to potentially higher profits from advertising. For example, if Walmart knows a user visited a site focused on infant health, it may choose to also advertise baby strollers to the same user. Making this information available to advertisers, though, can affect ad prices, since often the set of websites visited by a user conveys information about how valuable the user is as a target for advertising. That is, cookie matching can influence the prices of advertising in the marketplace, which may impact websites' revenues in different directions, as discussed shortly. In this paper, we will focus on the incentives for and against cookie-sharing, asking when websites will voluntarily agree to participate in cookie-sharing, and what effects on prices and profits cookie-sharing creates.

How cookie matching works. Sharing cookie information is done through a service called *cookie matching* that is currently offered by most online advertising exchanges. Cookie-matching means that when a user visits a website, the ad exchange scrambles the cookie that they have placed on the user's computer (using what is called a collision-resistant one-way hash function; i.e., a function that transforms each cookie to a unique identifying string in a way that the transformation cannot be reversed), and passes the scrambled cookie as an identification number for the cookie to the interested advertisers.[1] This cookie id, while not revealing the contents of the cookie, enables these advertisers to build a mapping between their own cookies and the ad exchange's cookie ids and discover if they have interacted with this user before (for example, if the user has visited their website before, or if they have advertised to the user before). The advertisers will then be able to decide how much they want to bid based on this information. This sharing also enables the advertisers (both the winner of the auction and also the losers) to identify the user in his/her future visits (unless the cookie is deleted in the meantime).

The obvious benefit of this mechanism is that it allows advertisers to target users that have previously shown interest in their products. This targeting increases the value of advertising to the advertisers, and some of this increased value will be passed to the publishers in the long run. The not-so-obvious drawback for the publishers, especially premium publishers, is *information leakage*: a publisher whose site is often visited by high-value users owns a particularly valuable piece of information about these users. Without cookie matching, since advertisers have a high value for displaying their ads to these users and the number of impressions on this site is limited, the prices can be driven up in the auction. However, by passing the scrambled cookie to the advertisers, the advertisers will be able to target the same user elsewhere on cheaper publishers. This, in effect, increases the supply of advertising opportunities, which can drive the auction prices down. In this hypothetical scenario, sharing information dilutes the value of the supply on the premium publishers, while increasing the value of other publishers' supply. In other words, the added value of information is leaked from the premium publisher to the others.

The goal of this paper is to explore the latter aspect of cookie-matching, i.e., when a publisher leaks valuable information through cookie-matching that could harm their revenue (while helping other publishers). This is important, since such scenarios would present an obstacle to the universal adoption of the cookie-matching technology: a publisher who is harmed by information leakage would not voluntarily adapt this technology in the absence of any side payment.

We start with a discussion of various effects of cookie-matching in ad auctions.

[1] See [Google 2013] for a more detailed explanation of the cookie matching service.

1.1. The impact of cookie-matching on auction revenue

Sharing information about the supply in an auction (here, the users) naturally impacts the outcome of the auction and the revenue it generates. We can divide the effects of providing data on users through cookie-matching on the advertising market into four major categories:

— **More efficient allocation**: Data (whether it is labeling the impressions with features of the user, or with identifiable information like the scrambled cookie) allows advertisers to evaluate impressions more accurately, thereby increasing the efficiency of the allocation.

— **Market fragmentation**: This increased efficiency can also lead to the fragmentation of the market and decreased revenue for the auctioneer. An example is when two advertisers are competing for an ad slot, but one advertiser is only interested in male users while the other is interested in females. In this scenario, providing any data that helps advertisers distinguish male and female users will lead to a more efficient allocation, but it can also lead to a lower revenue in a second-price auction.

— **Better interaction with the user**: Cookie matching allows the advertisers to know how many times they have seen a user before, thereby personalizing their ad creative each time and avoiding advertising too many times to the same user. The latter effect is called frequency-capping.

— **Data leakage**: By passing the cookie information to the advertiser, the publisher (and the ad exchange platform) run the risk that the advertiser takes advantage of this data elsewhere (on different publishers or even on different platforms) to decrease her cost at the cost of decreased revenue to the publisher.

The market fragmentation effect (and its positive counterpart, allocation efficiency) is essentially the reverse of the *bundling* problem that has been studied in the auction theory literature [see McAfee et al. 1989]. In this paper, our focus is on the information leakage effect: we first provide a model that demonstrates this information leakage effect in equilibrium, and then show that in a large class of models, information leakage does not cause disagreement about whether or not to cookie-match in equilibrium— that is, cookie-matching can either increase or decrease publishers' revenues, but the direction of the revenue change is the same for all publishers.

1.2. Overview

Our main contribution is a simple model that exhibits the information leakage effect in cookie matching. Intuition as well as anecdotal evidence suggests that this is a real effect with many practical implications, and therefore is a phenomenon that one would like to capture in the equilibrium of a simple model. To achieve this, we employ a model that avoids having to deal with the complexity of the first two effects by assuming a homogeneous set of advertisers— a set of advertisers whose relative valuations of different users are similar, or consistent (*i.e.*, the vectors describing each advertiser's valuation of different users are scalar multiples of each other). This is a realistic model in many situations where there is already enough background information even without cookie-matching (such as impression type, or user demographics) to divide the market into fine-enough segments, each with a homogeneous set of competing advertisers that value the same users highly.

Surprisingly, the analysis of a simple special case of such a model (presented in Section 2) shows that in equilibrium, there is always agreement between publishers in terms of whether they would like to share cookie information or not— either all publishers' revenues increase with cookie matching, or all publishers obtain lower revenue (which of these occurs depends, essentially, on whether frequency capping is revenue positive or revenue negative). We then extend this result to a fairly general model in Section 2. Essentially, the only major assumption of our model is the homogeneity of advertisers. The results, presented in Section 3, suggest that the information leakage effect might not be as serious a problem in practice as one might suspect. Finally, we provide a simple model with only two types of advertisers where this result does not continue to hold— *i.e.*, information leakage from cookie matching indeed has opposite effects on the revenues of different publishers.

1.3. Related work

Information sharing by competitors has long been studied by economists. Much of this literature has focused on auctions, especially starting with the seminal work in Milgrom and Weber [1982], and most recently in Abraham et al. [2013]. However, this literature focuses on information provided *to* competitors, rather than shared *by* competitors; there are only a few studies focused on incentives to share by competitors (e.g. Clarke [1983], Gal-Or [1985]) which conclude that firms will not voluntarily share information. Finally, there is an extensive literature on information-sharing in a cartel environment; see e.g. Teece [1994] and the references therein. Information sharing is often viewed as a sign of collusion on the principle that firms have no incentive to provide information except to produce a cartel. The focus of the cartel papers is not on the incentives to join an information-sharing system but instead on the use of such a system to fix prices.

There are also a number of recent papers on the role of information and targeting in advertising. Bergemann and Bonatti [2011] and Fu, Jordan, Mahdian, Nadav, Talgam-Cohen, and Vassilvitskii [Fu et al.] study the effect of introducing targeting information on the revenue of ad auctions. Emek et al. [2012] and Sheffet and Miltersen [2012] discuss the algorithmic question of designing revenue-optimal signaling schemes in second-price auctions. Babaioff et al. [2012] study optimal mechanisms for selling information.

2. MODEL

The online advertising market consists of three kinds of agents— users, publishers, or the websites visited by users, and the advertisers who want to display their ads to users on these websites. The price of an ad impression in this market will depend on how much information about the corresponding user is available to advertisers at the time of bidding, since advertisers have different values for advertising to different users. This means that whether or not cookie-matching is being used in the marketplace will affects the price of advertising, and this effect may be different on different websites. We will be interested in computing the equilibrium prices of impressions with and without cookie-matching to determine how cookie-matching impacts publishers' revenues, and in using this analysis to investigate when the information leakage phenomenon can arise in an advertising market with cookie-matching. We now present a formal model which allows addressing this question.

There is a set of publishers, or websites, $W = \{w_1, w_2, \ldots\}$, and a number of users visiting these websites (we use publisher and website interchangeably throughout the paper). A number of advertisers are interested in advertising to these users, some of whom are more likely to purchase the advertisers' products than others, and are therefore preferred by the advertisers as advertising targets.

User model. We model the fact that different users differ in their response to ads, and therefore have different values to advertisers, by saying that a user is either a *high* type or a *low* type. We denote the proportion of users of type t, where $t \in \{H, L\}$, by p_t. Note that in the most general model, user types do not have to be restricted to be binary (high and low), and there could be multiple types of users corresponding to the differing revenues that the advertiser expects to make from advertising to these users. For the majority of this paper, we focus on this simplified model with only two types of users, since this turns out to be a setting rich enough to capture the essence of the arguments. However, we note that this assumption is for simplicity only; we will briefly discuss how the result can be generalized to more than two types in Section 3.3.

We model user browsing behavior as follows. We assume that at each time period, the user (irrespective of her type) leaves the system (*i.e.*, stops browsing) with probability q.[2] If she continues to browse (which happens with the remaining probability $1 - q$), she chooses to visit website $w \in W$ with probability $p_{t,w}$, where this probability now depends on the user's type $t \in \{H, L\}$. We assume that these random choices of whether to continue, and which website to visit if continuing, are made independently at each step.

Advertiser model. There are N advertisers, each of whom has some positive value for a high-type user, and a value of zero for a low-type user. Advertisers cannot directly observe the type of a user— rather, they can only infer it from the user's behavior. The ith advertiser has a value of v_i for advertising to a high-type user for the first time; we assume the advertiser does not obtain any additional utility by advertising to the same user more than once. We number advertisers in decreasing order of values, so that $v_1 \geq v_2 \geq \cdots \geq v_N$. A special case of interest is that of a fully competitive market, *i.e.*, when the v_i's are all the same value v.

We use this model to study incentives for information sharing by comparing the expected revenue that publishers, or websites, obtain under two types of information sharing regimes— one with cookie-matching and one without. Without cookie-matching, any user appears to be a new user to every advertiser on every visit, since there is no information preserved between visits. When cookie-matching is available, a scrambled version of the user's cookie is sent to all advertisers every time she visits a website, so that when an advertiser bids for an impression, she knows the entire sequence of websites that the user u associated with that impression has visited so far. We analyze and compare the revenue in both models by computing market equilibrium prices and allocation, *i.e.*, a set of prices for impressions and an allocation of impressions to advertisers at which

— every impression for which at least one advertiser has non-zero value is sold, and
— each advertiser weakly prefers the impressions she receives to any other set of impressions.

Sometimes there can be more than one set of prices satisfying the above conditions. For example, if there is only one impression and two bidders with values v_1 and v_2 interested in this impression, then allocating the impression to the higher bidder at any price between v_1 and v_2 satisfies both of the above conditions. In such cases, we study the lowest-price market equilibrium (in this example, the equilibrium at price v_2), which is a natural generalization of the second price auction prices. Note that it is not a priori obvious that such a "lowest-price" equilibrium should exist: our proof also

[2]The assumption that the probability of leaving in each step is constant is consistent with the empirical observation that the number of websites visited during a session follows exponential decay; see, for example, [Ortega and Aguillo 2010].

establishes the existence of such an equilibrium (alternatively, one can apply the result of Demange, Gale, and Sotomayor [1986] that shows that such a canonical equilibrium exists). The revenue per impression of a website w_i (also called the revenue of publisher i) is the expected price of an impression on this website in such an equilibrium.

Our results are structured around a *homogeneity* assumption: we say that advertiser valuations are homogenous if the value of a user to each advertiser can be written as a product of an advertiser-specific value, and a user-type specific value (for example, zero for low-type users and one for high-type users). In other words, the relative valuations of users are consistent across advertisers. As we will see, the information leakage phenomenon that is intuitively expected with cookie matching in fact cannot arise when advertiser values have this property (as is the case in the model with high- and low-type users defined in this section)— all publishers' revenues change in the same direction (*i.e.*, they all increase or all decrease) as a result of cookie matching. As we will show in Section 4, however, when advertiser values do not have this consistency property, the heterogeneity in advertiser valuations can indeed lead to the information leakage phenomenon— we illustrate this via a simple example where a second group of advertisers that do not care about the type of the user (*i.e.*, with the same value for high and low-type users) are added to the market.

A special case. We use the following special case of the model as an illustration throughout the paper. There are two websites or publishers, w_1 and w_2, where w_1 is the 'premium' website (*i.e.*, the website with higher-value users) and w_2 is the non-premium website. A high-type user visits each website with probability $1/2$, whereas a low-type user only visits w_2— this means that visiting w_1 is a clear signal that the user is of high-type. Half the user population is high-type, while the other half is low-type, *i.e.*, $p_H = p_L = 1/2$. Also, all advertisers derive the same value from advertising to high-type users: $v_i = v$ for all i. This example represents a case where publisher w_1 has a valuable piece of information about its visitors (while w_2's information is not so valuable).

3. PUBLISHER REVENUE WITH HOMOGENEOUS ADVERTISERS

In this section, we analyze the expected revenue of each publisher in the models with and without cookie-matching when advertisers are homogeneous (*i.e.*, advertisers' values for users are all constant multiples of the same vector). The phenomenon of interest to us is disagreement between different publishers about whether or not to share information. More precisely, we would like to know if there are scenarios where providing cookie-matching increases the revenue of one publisher at the expense of another publisher. Our main result is the following theorem, which proves that in the model with homogeneous advertisers, this will never happen: all publishers agree about whether they would like to participate in cookie matching.

THEOREM 3.1. *When advertisers are homogenous, the expected revenue per impression of a website w, both with and without cookie-matching, is proportional to*

$$\beta_w := \frac{p_H p_{H,w}}{\sum_{t \in \{H,L\}} p_t p_{t,w}}.$$

Therefore, either for all websites w_i, the revenue per impression of w_i in the model with cookie-matching is greater than its revenue per impression in the model without cookie-matching, or the reverse inequality holds for all w_i.

Note that the quantity β_w is the fraction of impressions on w that are from high-type users. Therefore, the above theorem shows that in both models, all websites get the same expected revenue per high-type visitor.

The proof of Theorem 3.1 is presented in Sections 3.1 (for the model without cookie-matching) and 3.2 (for the model with cookie-matching). We discuss further generalizations of this result in Section 3.3, and finally numerically examine the calculated revenues in the case of the simplest model (with two websites and half the users being type H) in Section 3.4.

Intuition behind the proof. Our proof of Theorem 3.1 is based on calculating the equilibrium prices of impressions in both scenarios with and without cookie matching. With cookie matching, the price of an impression depends only on the user history, and not on the website on which the impression occured— more precisely, the price depends on the website only to the extent that it constitutes (the last) unit of information available to advertisers to update their belief about the type of the user. This means that for every history S of the websites the user has visited, we need to calculate an equilibrium price $\lambda(S)$ at this history. This is done by writing equations that capture the fact that prices are at an equilibrium. More specifically, the equilibrium equation in this case (Equation (5) in Section 3.2) says that the advertiser winning an impression (which is the advertiser with the highest bid who still has not displayed her ad to the user) is indifferent between winning the impression and waiting to win the next impression (if any) by the same user. This gives a recurrence relation that gives the equilibrium price at a history in terms of the equilibrium price at longer histories. This recurrence can be solved in combination with a terminating condition that says that when there is no remaining bidder, the price should drop to zero.

Without cookie matching, the price of an impression depends entirely on which website the impression occurs, since advertisers have no other information available to discern the user's type (so for example in the special case, prices will be higher on w_1 than w_2, since a user on w_1 is definitely a high-type user, whereas a user on w_2 can be either high or low-type). The equilibrium equation in this case will be in terms of variables θ_w representing the price of an impression on website w, as well as variables $x_{a,w}$, representing the fraction of the traffic of w that advertiser a purchases. The utility function of each advertiser can be written in terms of these variables, and first-order conditions of optimality of this utility function (in terms of the variables $x_{a,w}$ that are under a's control) give us the equilibrium equations. These equations give us the equilibrium prices and allow us to compare the publishers' revenues with and without cookie matching.

3.1. Analysis of the model without cookie-matching

In the model without any cookie-based user tracking, all impressions on a website look the same. Therefore, for each website w, there is a single user-independent price θ_w per impression. We now write equilibrium conditions for these prices.

Consider the utility maximization problem from the perspective of one fixed advertiser a, who needs to decide what fraction $x_{a,w}$ of traffic on each website w to buy. We compute the utility per user that this advertiser derives from a particular choice of allocation $x_{a,w}$. Fix a user of type $t \in \{H, L\}$. The expected total number of websites this user visits is $1/q$. On each such visit, the probability that she visits website w is $p_{t,w}$, in which case she sees the ad of a with probability $x_{a,w}$. Therefore, for a user of type t, a pays a total expected cost of

$$\frac{1}{q}\sum_{w\in W} p_{t,w}x_{a,w}\theta_w.$$

Also, the probability that the user continues to browse for exactly i steps is $q(1-q)^{i-1}$, and the probability of seeing a's ad on any such step where the user continues browsing and visits some website is $x_{a,t} := \sum_{w\in W} p_{t,w}x_{a,w}$. So the total probability of a particular user of type t getting exposed to a's ad at least once can be written as:

$$1-\sum_{i\geq 1} q(1-q)^{i-1}(1-x_{a,t})^i = \frac{x_{a,t}}{q+(1-q)x_{a,t}}.$$

Therefore, the total utility that a derives from a random user (of either type) is

$$U = p_H \cdot v_a \frac{x_{a,H}}{q+(1-q)x_{a,H}} - \frac{1}{q}\sum_{t\in\{H,L\}} p_t \sum_{w\in W} p_{t,w}x_{a,w}\theta_w.$$

To optimize the utility of the advertiser with respect to the fractions $x_{a,w}$ of the total inventory on each site that a buys, we need to take the derivative of the above expression with respect to each $x_{a,w}$ (recall that $x_{a,H} := \sum_{w\in W} p_{H,w}x_{a,w}$):

$$\frac{\partial U}{\partial x_{a,w}} = v_a p_H \frac{p_{H,w}q}{(q+(1-q)x_{a,H})^2} - \sum_{t\in\{H,L\}} \frac{p_t p_{t,w}\theta_w}{q}$$

The above derivative is zero if and only if

$$\theta_w = v_a p_H p_{H,w}(1+(\frac{1}{q}-1)x_{a,H})^{-2}(\sum_{t\in\{H,L\}} p_t p_{t,w})^{-1}$$

or

$$\theta_w = v_a(1+(\frac{1}{q}-1)x_{a,H})^{-2}\beta_w. \tag{1}$$

This means that for every advertiser a and website w, the allocations $x_{a,w}$ and prices θ_w are such that either

- Equation (1) holds; or
- $x_{a,w} = 0$ and $\theta_w > v_a(1+(\frac{1}{q}-1)x_{a,H})^{-2}\beta_w$; or
- $x_{a,w} = 1$ and $\theta_w < v_a(1+(\frac{1}{q}-1)x_{a,H})^{-2}\beta_w$.

We are now ready to complete the proof of Theorem 3.1 in the case of no cookie-matching. Assume, for contradiction, that at a lowest-price equilibrium, for two websites w and w', we have $\theta_w/\beta_w > \theta_{w'}/\beta_{w'}$. For any advertiser a, if $x_{a,w} > 0$ and $x_{a,w'} < 1$, we have

$$v_a(1+(\frac{1}{q}-1)x_{a,H})^{-2} \geq \theta_w/\beta_w > \theta_{w'}/\beta_{w'} \geq v_a(1+(\frac{1}{q}-1)x_{a,H})^{-2}, \tag{2}$$

which is a contradiction. Therefore, for every a, either $x_{a,w} = 0$ or $x_{a,w'} = 1$. Since at the equilibrium there must be at least one a^* with $x_{a^*,w} > 0$, for this a^*, we have $x_{a^*,w'} = 1$. This means that no other $a \neq a^*$ can have $x_{a,w'} = 1$, and therefore for all such a's, we have $x_{a,w} = 0$, implying that $x_{a^*,w} = 1$. Since this argument holds for every two w, w' with $\theta_w/\beta_w > \theta_{w'}/\beta_{w'}$, we conclude that a^* must have bought everything, *i.e.*, $x_{a^*,w} = 1$ for all w.

We now complete the proof using the fact that the equilibrium is a *lowest-price* equilibrium. Take a website w with the maximum value of θ_w/β_w. It is easy to see that if we slightly decrease the price of impressions at this website, no advertiser $a \neq a^*$ will still be interested in buying these impressions. Therefore we are still at an equilibrium, which contradicts the assumption on minimality of prices.

The contradiction shows that at a lowest-price equilibrium, for every two websites w and w', we must have $\theta_w/\beta_w = \theta_{w'}/\beta_{w'}$. $\square$

A closed form in the simple model. We can give a simple closed-form expression for the revenue in the case that all advertisers have the same value v for advertising to high-type users. In this case, due to symmetry, in the equilibrium we must have $x_{a,H} = 1/N$ for every advertiser a. Therefore, Equation (1) implies

$$\theta_w = v(1 + (\frac{1}{q} - 1)\frac{1}{N})^{-2}\beta_w. \tag{3}$$

Plugging in the parameters of the simple model, we get:

$$\theta_1 = (1 + \frac{1-q}{qN})^{-2}v \qquad \text{and} \qquad \theta_2 = (1 + \frac{1-q}{qN})^{-2}v/3.$$

3.2. Analysis of the model with cookie-matching

We now analyze the model in the cookie-matching regime. In this case, each impression comes with a complete history of the user (*i.e.*, the websites she has visited), and therefore the equilibrium price of advertising to the user can depend on this history. We denote the sequence of websites the user has visited by $S = w_{i_1}, w_{i_2}, \ldots, w_{i_k}$. The length of this history is denoted by $|S| = k$. We denote a history S followed by a visit to a website $w \in W$ by appending w to S as $S.w$. The price at the history S is denoted by $\lambda(S)$.

We write the equilibrium conditions for the price at a history S of length k. At this history, $k-1$ advertisers (that can be shown by induction to be the advertisers $1, \ldots, k-1$, *i.e.*, the advertisers with the top $k-1$ values) have already won an impression, and therefore only advertisers $k, \ldots, N$ are interested. The value of the ith advertiser for buying this impression is $\Pr[H|S].v_i - \lambda(S)$, where $\Pr[H|S]$ denotes the probability that the user is of type H, given the history S of the sites she has visited.

The utility of this advertiser for waiting is the probability that the user returns, which is $(1-q)$, times the utility conditioned on her return. If the user returns and visits a website w, the utility of the advertiser is $v_i - \lambda(S.w)$ if the user is of high type and $-\lambda(S.w)$ if she is of low type. Therefore, the overall utility of the advertiser if the user returns can be written as:

$$\begin{aligned}\Pr[H|S] \cdot \sum_{w \in W} p_{H,w} \cdot (v_i - \lambda(S.w)) + \Pr[L|S] \cdot \sum_{w \in W} p_{L,w} \cdot (-\lambda(S.w)) \\ = \Pr[H|S]v_i - \sum_{w \in W} \Pr[w|S]\lambda(S.w),\end{aligned} \tag{4}$$

where $\Pr[w|S] = \Pr[H|S].p_{H,w} + \Pr[L|S].p_{L,w}$ is the probability that a user visits the website w after the history S.

Therefore, the equilibrium condition says that for the advertiser who wins the present impression, the buy-now utility of $\Pr[H|S].v_i - \lambda(S)$ is greater than or equal to $(1-q)$ times the expression in (4), and for the other advertisers the reverse inequality holds. This implies that the advertiser winning this impression should be the advertiser k, and in order to get the lowest-price equilibrium, we must have equality for the advertiser $k+1$. Thus, the equilibrium condition implies:

$$\begin{aligned}\Pr[H|S].v_{k+1} - \lambda(S) \ &= \\ &(1-q)\left(\Pr[H|S]v_{k+1} - \sum_{w\in W} \Pr[w|S]\lambda(S.w)\right). \qquad (5)\end{aligned}$$

This implies the following recurrence that gives the price at any history in terms of prices at longer histories.

$$\lambda(S) \ = \ q.\Pr[H|S].v_{k+1} + (1-q)\sum_{w\in W}\Pr[w|S].\lambda(S.w). \qquad (6)$$

For the base of this recurrence, we have

$$\forall\ S, |S| \geq N: \quad \lambda(S) = 0. \qquad (7)$$

This is because after any history that contains at least N page visits, all but at most one of the advertisers have already advertised to the user and therefore the price drops to zero.

Using this recurrence, and induction on the length of the history, we obtain the following formula for the price at any given history:

$$\lambda(S) = \Pr[H|S].\sum_{i=k+1}^{N} q(1-q)^{i-k-1}v_i \qquad (8)$$

For each website $w \in W$, the expected revenue per user of this website can be written in terms of the prices $\lambda(.)$ as follows:

$$\text{Revenue per user of } w \ = \sum_{S=(w_{i_1},\ldots,w_{i_{k-1}},w)} \Pr[S]\lambda(S) \qquad (9)$$

Using (8) and (9), we can write the expected revenue per user of a website w as follows:

$$\text{Revenue per user of } w \ = \sum_{k=1}^{N}\sum_{S}\Pr[S].\Pr[H|S].\sum_{i=k+1}^{N} q(1-q)^{i-k-1}v_i,$$

where the second summation is over all histories S of length k that end in w. Using Bayes' rule, we have

$$\text{Revenue per user of } w \; = \; \sum_{k=1}^{N} \sum_{S} p_H . \Pr[S|H]. \sum_{i=k+1}^{N} q(1-q)^{i-k-1} v_i .$$

The summation $\sum_S \Pr[S|H]$ is equal to the probability that a high-type user visits at least k websites, and chooses the website w on her kth visit. This can be written as $(1-q)^{k-1} p_{H,w}$. Using this, we can simplify the above expression:

$$\begin{aligned}
\text{Revenue per user of } w \; &= \; \sum_{k=1}^{N} p_H .(1-q)^{k-1} p_{H,w} . \sum_{i=k+1}^{N} q(1-q)^{i-k-1} v_i \\
&= \; p_H p_{H,w} q \sum_{i=2}^{N} \sum_{k=1}^{i-1} (1-q)^{i-2} v_i \\
&= \; p_H p_{H,w} q \sum_{i=2}^{N} (i-1)(1-q)^{i-2} v_i \qquad (10)
\end{aligned}$$

The summation in the above expression is independent of w. This means that website w's revenue per user is proportional to $p_H p_{H,w}$. Also, note that a random user on each visit chooses w with probability $\sum_{t \in \{H,L\}} p_t p_{t,w}$. Therefore, the expected number of impressions that a user generates on w is proportional to $\sum_{t \in \{H,L\}} p_t p_{t,w}$. Thus, the expected revenue per impression on w is proportional to $p_H p_{H,w} / (\sum_{t \in \{H,L\}} p_t p_{t,w}) = \beta_w$. This completes the proof of Theorem 3.1 in the model with cookie-matching. □

A closed-form expression for the simple model. In the case that all advertisers have the same value v, the pricing solution (8) can be simplified to

$$\lambda(S) = v . \Pr[H|S]. \left(1 - (1-q)^{N-k}\right) . \qquad (11)$$

Also, Equation (10) can be simplified as follows:

$$\begin{aligned}
\text{Revenue per user of } w \; &= \; p_H p_{H,w} q v \sum_{i=0}^{N-1} i(1-q)^{i-1} \\
&= \; \frac{p_H p_{H,w} v}{q} (1 - N(1-q)^{N-1} + (N-1)(1-q)^{N}) \qquad (12)
\end{aligned}$$

In the simple model, a random user in expectation creates $1/q$ impressions, and each impression will be on w_1 with probability $1/4$ and on w_2 with probability $3/4$. Therefore, the expected revenue per impression in this model is $(1-N(1-q)^{N-1}+(N-1)(1-q)^{N})v$ for the website w_1, and $(1-N(1-q)^{N-1}+(N-1)(1-q)^{N})v/3$ for w_2.

3.3. Generalizing the model

Our results so far have been presented in a model where there are only two types of users. This assumption can be relaxed, as follows. Suppose the user can be any of the types in a set $\mathcal{T} = \{t_1, t_2, \ldots\}$. The fraction of the users of type t is p_t, and users of this type visit website w with probability $p_{t,w}$ in each stage. Denote the value of an advertiser a for a user of type t by $v_{a,t}$. We say that the advertisers are homogeneous if for every advertiser a and user type t, the value $v_{a,t}$ can be written as $\gamma_t v_a$, where v_a only depends on the advertiser and γ_t only depends on the user type — for example, γ_t could be the conversion rate (*i.e.*, probability of purchasing the product) of users of type t, and v_a the profit per conversion for advertiser a. The case of two types H, L corresponds to setting $\gamma_H = 1, \gamma_L = 0$.

In this section, we generalize Theorem 3.1 to this more general model. Note, however, that this is not a strict generalization: unlike Theorem 3.1, here we cannot prove that the equilibrium is the lowest-price equilibrium in the model without cookie-matching.

THEOREM 3.2. *With homogeneous advertisers and multiple user types as defined above, in both the models with and without cookie matching there is an equilibrium in which the expected revenue per impression of a website w is proportional to*

$$\beta_w := \frac{\sum_t p_t p_{t,w} \gamma_t}{\sum_t p_t p_{t,w}}.$$

In the model with cookie matching, this equilibrium is the unique lowest-price equilibrium. Therefore, in these equilibria, either the revenue per impression of w_i in the model with cookie-matching is greater than its revenue per impression in the model without cookie-matching for all websites w_i, or the reverse inequality holds for all w_i.

PROOF. We start with the model without cookie matching. Similar to the analysis in Section 3.1, we let θ_w denote the price per impression on website w and $x_{a,w}$ denote the fraction of the traffic of website w that advertiser a buys, and write optimality conditions for these allocations and prices. The expected total cost a pays for a fixed user of type $t \in \mathcal{T}$ is

$$\frac{1}{q} \sum_{w \in W} p_{t,w} x_{a,w} \theta_w.$$

Also, the probability that a particular user of type t getting exposed to a's ad at least once can be written as:

$$1 - \sum_{i \geq 1} q(1-q)^{i-1}(1 - x_{a,t})^i = \frac{x_{a,t}}{q + (1-q)x_{a,t}},$$

where $x_{a,t} := \sum_{w \in W} p_{t,w} x_{a,w}$. Therefore, the total utility that a derives from a random user is

$$U = \sum_{t \in \mathcal{T}} p_t \gamma_t v_a \frac{x_{a,t}}{q + (1-q)x_{a,t}} - \frac{1}{q} \sum_{t \in \mathcal{T}} p_t \sum_{w \in W} p_{t,w} x_{a,w} \theta_w.$$

The derivative of the above expression with respect to the fractions $x_{a,w}$ can be written as:

$$\frac{\partial U}{\partial x_{a,w}} = \sum_{t \in \mathcal{T}} p_t \gamma_t v_a \frac{p_{t,w} q}{(q + (1-q)x_{a,t})^2} - \sum_{t \in \mathcal{T}} \frac{p_t p_{t,w} \theta_w}{q}$$

The above derivative is zero if and only if

$$\theta_w = \left(\sum_{t \in \mathcal{T}} p_t \gamma_t v_a p_{t,w} (1 + (\frac{1}{q} - 1) x_{a,t})^{-2} \right) (\sum_{t \in \mathcal{T}} p_t p_{t,w})^{-1}.$$

Let $f(x) = (1 + (\frac{1}{q} - 1)x)^{-2}$. Note that $f : [0,1] \mapsto [q,1]$ is a continuous decreasing function mapping 0 to 1 and 1 to q. Using this notation, the above equation can be written as:

$$\theta_w = \left(\sum_{t \in \mathcal{T}} p_t \gamma_t p_{t,w} f(x_{a,t}) v_a \right) (\sum_{t \in \mathcal{T}} p_t p_{t,w})^{-1}. \tag{13}$$

We now give an algorithm for constructing an allocation and a set of prices that are in equilibrium. For every w, the algorithm maintains $\theta_w = \delta\beta_w$ for a value δ that will be adjusted until an equilibrium is reached. Furthermore, for every advertiser a, the algorithm maintains the property that $x_{a,w}$ is the same for all websites w. We denote this common value by x_a. Clearly, this means that $x_{a,t} = x_a$ for all types t.

Initially, we set $x_a = 0$ for all advertisers a. Throughout the algorithm, we maintain

$$\delta = \max_a (f(x_a) v_a), \tag{14}$$

which initially means $\delta = \max_a v_a$. We continuously decrease δ, and to maintain the invariant (14), we increase the value of x_a for a's which achieve the maximum of $f(x_a)v_a$. We stop as soon as $\sum_a x_a$ reaches 1.

First, note that since the algorithm keeps increasing $\sum_a x_a$, it will eventually stop. At the point that the algorithm stops all of the following conditions are satisfied:

(i) For all w, $\theta_w = \delta\beta_w$.
(ii) For all a, $\delta \geq f(x_a)v_a$.
(iii) For all a with $x_a > 0$, we have $f(x_a)v_a = \delta$.
(iv) $\sum_a x_a = 1$.

By properties (i) and (iii), for every a with $x_a > 0$ and every w, the equality (13) holds, and therefore the derivative of U with respect to $x_{a,w}$ is zero. Also, for every a, by (i) and (ii), the right-hand side of (13) is less than or equal to its left-hand side, which means that the derivative of U with respect to $x_{a,w}$ is non-positive. Finally, U is a sum of linear and inverse-linear functions (with negative coefficients), and is therefore a concave function of $x_{a,w}$'s. Therefore, the allocation computed by the algorithm maximizes the utility function U. Since by property (iv) the total demand at this allocation equals the total supply, this set of allocation and prices forms an equilibrium. It is clear that in this equilibrium, the revenue per impression of a website w is θ_w, which is proportional to β_w. This completes the proof of the theorem in the model without cookie matching.

We now analyze the model with cookie matching. As in the analysis in Section 3.2, the equilibrium price of an impression by a user can depend on the user's browsing history S, and is denoted by $\lambda(S)$. At such a history S with $|S| = k$, the advertisers $1, \ldots, k-1$ have already won an impression, and therefore only advertisers $k, \ldots, N$ are interested. The value of the ith advertiser for buying this impression is $\mathrm{E}[\gamma_t|S].v_i - \lambda(S)$, where $\mathrm{E}[\gamma_t|S]$ denotes the expectation of the value of γ_t, where t is the type of the user, given the history S of the sites she has visited.

The utility of this advertiser for waiting is the probability that the user returns, which is $(1 - q)$, times the utility conditioned on her return. If the user returns and visits a website w, the utility of the advertiser is $\mathrm{E}[\gamma_t|S.w].v_i - \lambda(S.w)$. Therefore, the expected utility of the advertiser, if the user returns, can be written as

$$\begin{aligned} &\sum_{t \in \mathcal{T}} \Pr[t|S] \cdot \sum_{w \in W} p_{t,w} \cdot (\gamma_t.v_i - \lambda(S.w)) \\ &\qquad = \mathbf{E}[\gamma_t|S]v_i - \sum_{w \in W} \Pr[w|S]\lambda(S.w), \end{aligned} \tag{15}$$

where $\Pr[w|S] = \sum_{t\in\mathcal{T}} \Pr[t|S].p_{t,w}$ is the probability that a user visits the website w after the history S. As in Section 3.2, at a lowest-price equilibrium, the expected utility of buying the next impression for advertiser $k+1$ should be the same as this advertiser's expected utility for waiting. This translates to the following equation:

$$\begin{aligned}\mathbf{E}[\gamma_t|S].v_{k+1} - \lambda(S) \ =& \\ (1-q)\left(\mathbf{E}[\gamma_t|S]v_{k+1} - \sum_{w\in W}\Pr[w|S]\lambda(S.w)\right).&\end{aligned} \tag{16}$$

This yields a recurrence that gives the price at any history in terms of prices at longer histories. As in Section 3.2, by induction, the following closed form solution can be proved for this recurrence (together with the initial condition $\lambda(S) = 0$ for $|S| \geq N$):

$$\lambda(S) = \mathbf{E}[\gamma_t|S]. \sum_{i=k+1}^{N} q(1-q)^{i-k-1}v_i \tag{17}$$

For each website $w \in W$, the expected revenue per user of this website can be written as follows:

$$\begin{aligned}\text{Revenue per user of } w \ &= \sum_k \sum_{S=(w_{i_1},\ldots,w_{i_{k-1}},w)} \Pr[S]\lambda(S) \\ &= \sum_{k=1}^{N} \sum_{S=(w_{i_1},\ldots,w_{i_{k-1}},w)} \Pr[S].\mathbf{E}[\gamma_t|S]. \sum_{i=k+1}^{N} q(1-q)^{i-k-1}v_i.\end{aligned}$$

By Bayes' rule,

$$\begin{aligned}\sum_{S=(w_{i_1},\ldots,w_{i_{k-1}},w)} \Pr[S].\mathbf{E}[\gamma_t|S] \ &= \sum_{t\in\mathcal{T}} \gamma_t \sum_{S=(w_{i_1},\ldots,w_{i_{k-1}},w)} \Pr[S].\Pr[t|S] \\ &= \sum_{t\in\mathcal{T}} \gamma_t p_t \sum_{S=(w_{i_1},\ldots,w_{i_{k-1}},w)} \Pr[S|t] \\ &= \sum_{t\in\mathcal{T}} \gamma_t p_t (1-q)^{k-1} p_{t,w}.\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Revenue per user of } w \ &= \sum_{k=1}^{N} \sum_{t\in\mathcal{T}} \gamma_t p_t (1-q)^{k-1} p_{t,w}. \sum_{i=k+1}^{N} q(1-q)^{i-k-1}v_i, \\ &= \sum_{t\in\mathcal{T}} \gamma_t p_t p_{t,w} q \sum_{i=2}^{N} \sum_{k=1}^{i-1} (1-q)^{i-2} v_i \\ &= \sum_{t\in\mathcal{T}} \gamma_t p_t p_{t,w} q \sum_{i=2}^{N} (i-1)(1-q)^{i-2} v_i\end{aligned} \tag{18}$$

The summation in the above expression is independent of w. This means that website w's revenue per user is proportional to $\sum_{t\in\mathcal{T}} \gamma_t p_t p_{t,w}$. Also, note that a random user

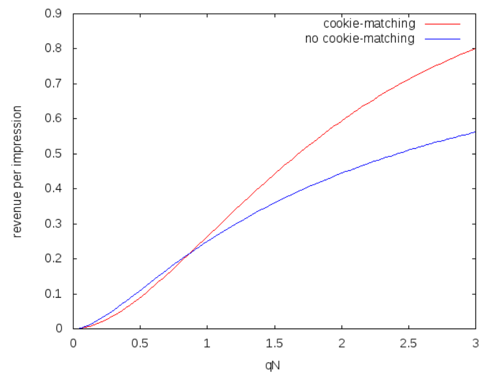

Fig. 1. Revenue comparison between the two scenarios in the simple model

on each visit chooses w with probability $\sum_{t\in\mathcal{T}} p_t p_{t,w}$. Therefore, the expected number of impressions that a user generates on w is proportional to $\sum_{t\in\mathcal{T}} p_t p_{t,w}$. Thus, the expected revenue per impression on w is proportional to $\sum_{t\in\mathcal{T}} \gamma_t p_t p_{t,w} / (\sum_{t\in\mathcal{T}} p_t p_{t,w}) = \beta_w$. This completes the proof of Theorem 3.2 in the model with cookie-matching. $\square$

3.4. Numerical examination of the simple model

It is instructive to look at the closed-form expressions for the revenue in the simple model. Setting $v = 1$ and N to be a large number, we plot the revenue per impression of w_1 as a function of qN in Figure 1. As can be seen in this figure, there is a range of parameters for which the model without cookie-matching achieves a higher revenue for publishers than the model with cookie-matching. For example, this happens when q is roughly $1/(2N)$, which means that each user visits about $2N$ websites before quitting. The intuitive reason is that in this range, the supply in the cookie matching model ($2N$ impressions per user) is more than the demand (N advertisers), leading to a low price. In the model without cookie-matching, the inefficiency due to the possibility of one advertiser advertising multiple times to the same user artificially decreases the supply, thereby increasing the prices. As the supply ($1/q$ impressions per user) gets more in line with demand (N), the cookie-matching model yields higher revenue for all publishers than the model without cookie-matching.

4. A SCENARIO WITH INFORMATION LEAKAGE

In this section, we show that when advertisers are heterogenous, the impact of cookie matching on different publishers can indeed be different, leading to a loss of revenue for some publishers and an increase in revenue for others. This illustrates that the results on publishers agreeing about whether or not to cookie-match depends on the homogeneity of advertisers' valuations of users.

Consider the simple model where there are N advertisers, whom we will call the type-A advertisers, who value high type users at v and low type users at 0. Suppose in addition that there are N_B type B advertisers that have a value of R the first time their ad is shown, *regardless* of the user type. Type-A advertisers represent advertisers who are able to use the cookie information to better target their ads to the users, whereas type-B's are generic advertisers that are willing to mass advertise to everyone.

We will show that in this setting, there exists a setting of $N, N_B, R, v,$ and q, so that w_1 has higher expected revenue (at equilibrium) in the setting without cookie matching, whereas w_2 has higher expected revenue in the setting with cookie matching. This is precisely the information leakage scenario where the premium website w_1 suffers a loss in revenue due to a dilution of its supply of high valued users.

4.1. Equilibrium revenue with no cookie matching

From the point of view of type A advertisers, the equilibrium conditions are the same as the ones in Section 3.1. Now consider advertisers of type B. If $N_B \gg \frac{1}{q}$ then advertisers of type B will be willing to pay $R - \epsilon$ per impression, as the chances of a single advertiser seeing the same user twice are nearly 0.

If the prices θ_1, θ_2 and the value of R is such that $\theta_1 > R > \theta_2$ then advertisers of type A will never be allocated any users from w_2. In this case, $x_{a,H} = \frac{1}{2N}$, as the N type A advertisers evenly split all of the high valued impressions coming to w_1. By Equation (1), the prices that support such an allocation are: $\theta_1 = v\left(1 + \frac{1-q}{2qN}\right)^{-2}$. Also, at any price greater than $\theta_2 = \theta_1/3$, type A advertisers do not want any of the impressions on w_2.

We can now compute the revenue on each website.

LEMMA 4.1. *Let $\theta_1 = v\left(1 + \frac{1-q}{2qN}\right)^{-2} = 3\theta_2$. If $\theta_1 > R > \theta_2$, then the revenue per impression to w_1 is exactly θ_1 and the revenue per impression to w_2 is $R - \epsilon$ for an ϵ that tends to zero as N_B tends to infinity.*

4.2. Cookie matching

When cookie matching is enabled, then the price advertisers are willing to pay is determined by the previous history of the user. In particular, any user who has ever visited w_1 is guaranteed to be a type H user, regardless of which website he is currently visiting. On the other hand, if the history S represents k visits to w_2 (and no visit to w_1), the probability $\Pr[H|S]$ that the user with this history is a high-type can be calculated using the Bayes rule as follows:

$$\Pr[H|S] = \frac{\Pr[S|H] \cdot \Pr[H]}{\Pr[S|H] \cdot \Pr[H] + \Pr[S|L] \cdot \Pr[L]} = \frac{2^{-k} \cdot \frac{1}{2}}{2^{-k} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}} = \frac{1}{2^k + 1}$$

Therefore, if

$$\frac{v}{2^N + 1} > R, \tag{19}$$

then the expected value of a user who has visited w_2 k times without visiting w_1 to a type A advertiser is $v \cdot Pr[H|S] = \frac{v}{2^k+1} \geq R$. Therefore the expected value to a type A advertiser is larger than the value to a type B advertiser. This implies that the per impression price for visits $N+1$ and onwards for *any* user will be R.

Let λ_i be the price of the i-th impression of the user with a history containing at least one visit to w_1. Adapting Equation (5) to this specific setting, with the base case of the recurrence as $\lambda_j = R$ for all $j \geq N$, we get:

$$\lambda_k = v - (1-q)^{N-k}(v - R) \tag{20}$$

Therefore, the expected revenue per user to w_1 is:

$$\sum_{k=1}^{N} \frac{(1-q)^{k-1}}{4} \big(v - (1-q)^{N-k}(v-R)\big) + \sum_{k=N+1}^{\infty} \frac{(1-q)^{k-1}}{4} R.$$

Since each user creates $1/(4q)$ impressions on w_1 in expectation, the per impression revenue of w_1 is:

$$\begin{aligned}\theta_1' &= 4q \cdot \sum_{k=1}^{N} \frac{(1-q)^{k-1}}{4} \big(v - (1-q)^{N-k}(v-R)\big) + 4q \cdot \sum_{k=N+1}^{\infty} \frac{(1-q)^{k-1}}{4} R \\ &= v(1-(1-q)^N) - Nq(v-R)(1-q)^{N-1} + R(1-q)^N\end{aligned}$$

Moreover, the per impression revenue of w_2 strictly increases, since in addition to type B advertisers, type A advertisers also sometimes bid on impressions on w_2, and whenever they do so, their bid is strictly greater than R.

To demonstrate information leakage, we need to find a setting of R, v, N, q such that satisfy:

$$\begin{aligned}\theta_1 = v\left(1 + \frac{1-q}{2qN}\right)^{-2} &> R > v\left(1 + \frac{1-q}{2qN}\right)^{-2}/3 = \theta_2 \\ \frac{v}{2^{N+1}} &> R \\ \theta_1 = v\left(1 + \frac{1-q}{2qN}\right)^{-2} &> v(1-(1-q)^N) - Nq(v-R)(1-q)^{N-1} + R(1-q)^N = \theta_1'\end{aligned}$$

Setting $R = 0.03, v = 1, N = 4$ and $q = 0.05$ entails $\theta_1 = 0.0878$, $\theta_2 = 0.0293$ and $\theta_1' = 0.0436$, satisfying the three conditions above. This leads to a lower per impression revenue to the owner of w_1, and a higher per impression revenue to the owner of w_2 in the cookie matching case.

5. CONCLUSION

Cookie-matching is now commonplace on the internet, with many publishers, or websites, sharing cookie information with each other (for example, see [Perlich and Dalessandro 2013] or [Delo 2013]). We investigated publishers' incentives to share cookies and found, surprisingly, that when advertisers value users homogeneously, publishers agree about whether or not to share cookies— that is, either all publishers want to share cookie information, or no publishers want to share cookies.

This result can be understood as follows. In both scenarios (cookie-matching or no cookie-matching), advertisers are paying the expected value of advertising to a user. With cookie-matching, this expected value is contingent on the user's history of websites visited. Without cookie-matching, the expectation is taken only over the population of users visiting a website. Either way, cookie-matching does not change the *nature* of visitors to any website, but rather only what is *known* about them: this increase in knowledge either increases advertiser values— in which case all publishers unanimously agree that cookie-matching enhances revenues— or lowers advertiser values, in which case publishers prefer not to match. (Advertiser values can either rise or fall with cookie matching: it can rise because cookie matching improves the interaction with the user (e.g., permitting frequency capping), increasing value, or fall because it identifies a greater supply, decreasing values.)

When advertisers disagree about the relative value of users, however, cookie-matching can indeed cause the information leakage phenomenon, leading to a disagreement between publishers about whether or not to participate in cookie-matching. The simplest example illustrating this phenomenon involves two sites and two types of users. Site 1 attracts only type H users, while site 2 attracts type H and type L. Suppose some advertisers value only type H users, while the others are indifferent. With no cookie matching, advertisers buying on site 2 must advertise to both types of users, reducing the willingness to pay of advertisers who only value type-H users, so that impressions on this website are won by advertisers who are indifferent. With cookie matching, site 2 can sell some of the type-H impressions to advertisers that value this type highly. This increases the revenue of 2, since it increases the price of some impressions, while keeping the price of the remaining impressions almost intact. It also increases the supply of known type-H impressions, reducing the demand to site 1. Thus with cookie matching, site 1 loses revenue, and site 2 gains.

Further directions. There is much more to explore in cookie-matching. In particular, we have set aside the bundling aspects of the absence of cookie-matching; cookie-matching leads to market fragmentation. This leads to increased efficiency, but it can also lead to decreased revenue in a thin market. Quantifying the value of cookie matching to each publisher is an interesting open direction. Also, in cases such as the data leakage scenario presented in this paper, side-payments might overcome the conflict of interest among the publishers – what mechanisms can be devised to allow the publishers that benefit from cookie-matching to compensate the losing publishers and *buy* their consent to cookie-matching? Finally, understanding the user side of the cookie-matching game, and in particular the privacy concerns of the user is worthy of further study.

REFERENCES

ABRAHAM, I., ATHEY, S., BABAIOFF, M., AND GRUBB, M. 2013. Peaches, lemons, and cookies: Designing auction markets with dispersed information. In *Proceedings of the 14th ACM Conference on Electronic Commerce (EC'13)*. ACM, 7–8.

BABAIOFF, M., KLEINBERG, R., AND PAES LEME, R. 2012. Optimal mechanisms for selling information. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC'12)*. ACM, 92–109.

BERGEMANN, D. AND BONATTI, A. 2011. Targeting in advertising markets: Implications for offline vs. online media. *RAND Journal of Economics 42*, 414–443.

CLARKE, R. N. 1983. Collusion and the incentives for information sharing. *The Bell Journal of Economics 14,* 2.

DELO, C. 2013. Facebook inks deal with acxiom, epsilon, and others to show ads based on shopping habits. `http://adage.com/article/digital/facebook-announces-plan-show-ads-based-offline-shopping-habits/240054/`.

DEMANGE, G., GALE, D., AND SOTOMAYOR, M. 1986. Multi-item auctions. *Journal of Political Economy 94,* 4, 863–72.

EMEK, Y., FELDMAN, M., GAMZU, I., PAES LEME, R., AND TENNENHOLTZ, M. 2012. Signaling schemes for revenue maximization. In *Proceedings of the 13th ACM Conference on Electronics Commerce (EC'12)*.

FU, H., JORDAN, P., MAHDIAN, M., NADAV, U., TALGAM-COHEN, I., AND VASSILVITSKII, S. Ad auctions with data. In *Proceedings of the 5th International Symposium on Algorithmic Game Theory (SAGT)*.

GAL-OR, E. 1985. Information sharing in oligopoly. *Econometrica 53,* 2.

GOOGLE. 2013. Cookie matching. `https://developers.google.com/ad-exchange/rtb/cookie-guide`.

MCAFEE, R. P., MCMILLAN, J., AND WHINSTON, M. 1989. Commodity bundling by a monopolist. *Quarterly Journal of Economics*, 371–383.

MILGROM, P. AND WEBER, R. J. 1982. A theory of auctions and competitive biddin. *Econometrica 50,* 5, 1089–1122.

ORTEGA, J. L. AND AGUILLO, I. F. 2010. Differences between web sessions according to the origin of their visits. *Journal of Informetrics 4,* 3, 331–337.

PERLICH, C. AND DALESSANDRO, B. 2013. The data behind retargeting. `http://blog.ad-tech.com/the-data-behind-retargeting/`.

SHEFFET, O. AND MILTERSEN, P. B. 2012. Send mixed signals - earn more, work less. In *Proceedings of the 13th ACM Conference on Electronics Commerce (EC'12)*.

TEECE, D. J. 1994. Information sharing, innovation, and antitrust. *Antitrust Law Journal 62,* 2.