Online advertising and privacy *

Alexandre de Cornière[†], Romain de Nijs[‡]

Preliminary and incomplete

June 14, 2011

Abstract

We study a model in which an online platform makes a profit by auctioning an advertising slot that appears whenever a consumer visits its website. Several firms compete in the auction, and consumers differ in their preferences. Prior to the auction, the platform gathers data which is statistically correlated with consumers' tastes. We study whether it is profitable for the publisher to allow potential advertisers to access the data about consumers' characteristics before they bid. As in Ganuza (2004), the paltform's trade-off is between maximizing the willingness to bid and extracting bidders' rent. We identify a new trade-off, namely that the disclosure of information leads to a better matching between firms and consumers, but results in a higher equilibrium price on the product market. We find that the equilibrium price is an increasing function of the number of firms. As the number of firms becomes large, it is always optimal for the platform to disclose the information, but this need not be efficient, because of the distorsion caused by the higher prices.

*For useful discussions and suggestions we thank Bernard Caillaud, Chris Dellarocas, Gabrielle Demange, Nabil Kazi-Tani, Frédéric Koessler, Philippe Jéhiel, and seminar participants and Crest, the Paris School of Economics, the sixth bi-annual "Conference on The Economics of Intellectual Property, Software and the Internet" in Toulouse, the second "Workshop on the Economics of ICT" in Universidade de Evora, the 2011 IIOC conference

[†]Paris School of Economics and CREST (LEI). Email: adecorniere@gmail.com

[‡]Paris School of Economics and CREST (LEI). Email: romaindenijs@gmail.com

1 Introduction

The online advertising industry has been growing rapidly in the last decade, thanks, on the one hand, to the growth of the number of Internet users, and, on the other hand, to technological advances.[1] These advances have concerned the ability of firms to gather, stock and analyze considerable amount of data, but also their ability to use this data at a very high speed, making it possible to customize every interaction.

There are three main types of online advertising. Search advertising is the most important one, accounting for 43% of the total revenue from advertising.[2] In search advertising, firms bid for different keywords, so that the link to their website appears alongside the results on a search engine page. The second type of online advertising -third in importance- which represents around 14% of revenue, is classified advertising. This form of advertising is used a lot for recruitment offers, and directories of service providers.[3] Display advertising is the second most important type online advertising, with 33% of total revenue, and it is the one on which we focus in this paper. Let us describe briefly the actors involved in online display advertising: at opposite ends of the spectrum are the advertisers and the consumers, the former trying to reach the latter. Consumers visit various websites (publishers), such as nytimes.com or ESPN.go.com. Publishers typically have advertising space that they wish to sell, either directly with their sales team, or through an intermediary, for instance an advertising network (Google's Double Click, or Yahoo!Network). These networks aggregate supply (the publishers' side) and demand (the advertisers' side) for advertising space and play the role of match makers. Sometimes the advertising network is integrated with the publisher, Facebook or Google being the most prominent examples. The functioning of online advertising intermediated by an ad network is described in Figure 1. In this figure, user 2 visits publisher 1, user 3 visits publisher 2, and user 1 visits both publishers. Publishers register with the advertising network, which is in charge of filling the ad space on their websites. Advertisers 1 to n submit bids which may depend on users' characteristics as well as publishers' characteristics. In the example, users 1 and 2 see a different advertisement when they visit publisher 1. User 3 and 1 see the same advertisement when they visit publisher 2, but this advertisement is different from the one that user 1 sees on publisher 1's website.

Because of the technology advances mentioned above, the match making activity has experienced enormous improvements: not only is it possible to match every consumer to a different advertisement using real time auctions, but the accuracy of such a matching may be enhanced by the considerable amount of data that publishers and advertising networks have about consumers. The match may be based for instance on the correspondence between the publisher's

[1] See Evans (2008) and Evans (2009) for insightful discussions about this industry.

[2] source: International Advertising Bureau, quoted in Evans (2009)

[3] See Office Of Fair Trading (2010).

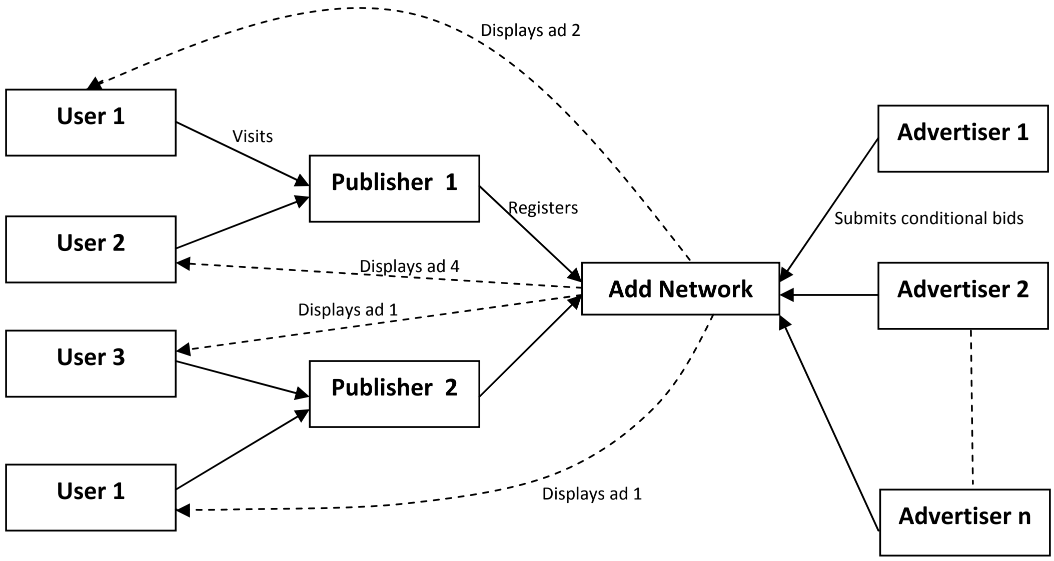

Figure 1: Non-integrated publishers and advertising network

website content and the advertisement, but also on data about the location of the consumer (obtained through the IP address), his past browsing history (obtained through cookies) or whatever information he gave to the publisher or its partners (through subscription questionnaires for instance, or any information left on his Facebook wall). These new opportunities may give firms additional incentives to acquire and use personal information about consumers, which has led regulators and consumers to express worries, or at least to acknowledge some potential pitfalls. Among these are privacy breaches or fraudulent use of personal information, but also practices of behavioral targeting and pricing.

In this paper, our focus is on the incentives of intermediaries such as advertising networks or integrated publishers to use the information they have gathered about consumers in order to increase their revenue from advertising. Do such practices have social value? Who benefits most from them?

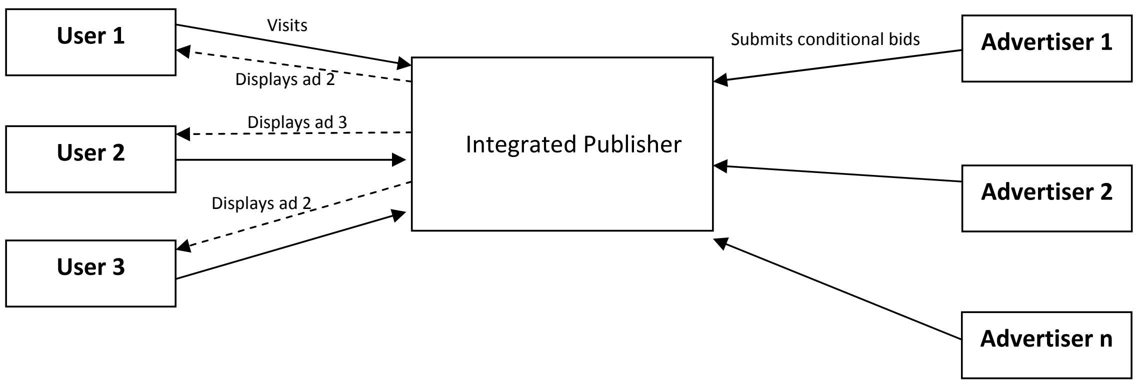

Figure 2: Integrated publisher

More specifically, the situation that we have in mind is the following (see Figure 2): a large number of web users (consumers) visit a website (a publisher), which makes profit by selling a single advertising slot through an auction. The website is integrated with an advertising network as is the case for Facebook, which allocates the slot to an advertiser. Users are heterogeneous in the sense that they do not derive the same value from consuming advertisers' products.

Thanks to its technology, the publisher gathers, for each consumer, information correlated to the consumer's willingness to pay for any product. The publisher does not how to interpret the information in terms of implied willingness to pay for different products, but advertisers are able to do it. For instance, the publisher knows that the consumer is a young male living in a metropolitan area, but it is not able to infer his willingness to pay for good A or B. On the other hand, firm A knows that young males living in a metropolitan area are especially likely to have a high willingness to pay for its product, whereas firm B offers a product which is less likely to be a good match for such consumers.

The heart of the problem for the publisher is the following: is it profitable to let advertisers learn consumers' characteristics ? The advantage of such a practice is that it increases the willingness to pay of the winning bidder with respect to the situation in which firms have no information about consumers. The potential drawback is that breaking the initial informational symmetry gives the winning bidder a positive informational rent. This trade-off appears also in Ganuza (2004). In our paper however, there is another effect which modifies the trade-off, namely the fact that the disclosure policy influences firms' pricing, so that disclosure induces firms to charge a higher price to consumers, since they know that the consumers they will face have on average a higher willingness to pay. Thus, roughly speaking, allowing firms to learn more about consumers enables them to increase their per-interaction revenue, but comes at the price of giving them a rent.

If the publisher decides to let firms learn consumers' characteristics, consumers experience two opposite effects: on the one hand, they are more likely to see a relevant advertisement, but on the other hand the prices that they face are higher.

We show that the publisher's and consumers' relative preferences between privacy and disclosure are not systematically opposed nor aligned. The publisher may provide too little or too much privacy, but in some cases it provides the right amount.

The paper is organized as follows: in section 2 we expose the model that illustrates the different trade-offs in a rather general way. In section 3 we caracterize symmetric equilibria under privacy and disclosure, and compare the two. In section 4, we study a simple example in order to better assess the desirability of privacy or disclosure from a social welfare standpoint. Section 5 discusses the related literature. [4]

[4]We plan on adding a section that extends the binary model to any number of firms, and we have several extensions that deal with targeted pricing, competition between platforms. We have not had the time to write them yet...

2 Model

The market we model is the following: there are n advertisers, who compete for a single slot on a publisher's website. There is a continuum of consumers who visit the website. A consumer's type is a vector $\theta = (\theta_1, ..., \theta_n)$. The $\theta_i$are independent and identically distributed according to a continuous cdf F over an interval set $[0, \overline{\theta}]$.The probability distribution function of θ_i is f.

If a consumer of type θ is matched with firm i, which sets a price p_i, firm i's profit is $\pi_i(p_i, \theta_i)$. Welfare and consumer's surplus, if consumer θ is matched with firm i, are noted respectively $W(p_i, \theta_i)$ and $S(p_i, \theta_i)$. We make the following set of assumptions:

Assumption 1 *For every i, π_i is twice continuously differentiable in both arguments.*

There exists $\overline{p}$ such that for all θ_i, and for all $p_i \geq \overline{p}$, $\pi_i(p_i, \theta_i) = 0$

Assumption 2 *For every price $p < \overline{p}$, $\pi_i(p, \theta_i) \geq \pi_j(p, \theta_j) \iff \theta_i \geq \theta_j$*

Assumption 3 *π_i is strictly concave in p_i over $[0; \overline{p}]$. For every θ_i, there exists $p^*(\theta_i) \in [0; \overline{p}]$ such that $\frac{\partial \pi_i(p^*(\theta_i), \theta_i)}{\partial p_i} = 0$.*

Assumption 4 *The profit function exhibits increasing differences : $\frac{\partial^2 \pi_i}{\partial p_i \partial \theta_i} \geq 0$*

Assumption 5 $\frac{\partial W}{\partial p_i} \leq 0, \quad \frac{\partial W}{\partial \theta_i} \geq 0, \quad \frac{\partial S}{\partial p_i} \leq 0, \quad \frac{\partial S}{\partial \theta_i} \geq 0$

Comments about the assumptions The main assumption among Assumptions 1 - 5 is probably Assumption 2. It implies that when $\theta_i > \theta_i'$, the demand curve conditional on θ_i is above the demand curve conditional on θ_i'. It is not satisfied when θ_i indexes a family of demand rotations (see Johnson and Myatt (2006)).

Assumptions 1 and 3 are made mainly for analytical convenience. In particular, $\overline{p}$ should be interpreted as the minimal price such that no consumer would be willing to buy a single unit.

An important corollary of Assumptions 3 and 4 is the following:

Lemma 1 *$p^*(\theta_i)$ is non-decreasing in θ_i.*

Proof : The proof is a classical result of monotone comparative statics (see Vives (2001) for instance). Let $\theta_i > \theta_i'$, and $p' > p^*(\theta_i)$. From Assumption 4 we have $\pi_i(p', \theta_i) - \pi_i(p^*(\theta_i), \theta_i) \geq \pi_i(p', \theta_i') - \pi_i(p^*(\theta_i), \theta_i')$. But, by Assumption 3, $\pi_i(p', \theta_i) - \pi_i(p^*(\theta_i), \theta_i) < 0$. Therefore p' cannot maximize $\pi_i(p, \theta_i')$. $\square$

We assume that the realization of θ is consumer's private information, but the platform observes a signal about θ. The platform does not know the mapping from the signal to the actual value of θ. It can choose to publicly reveal the value of the signal to advertisers. In that case, each firm i privately learns the value of θ_i.

One can imagine that θ_i is the score that firm i would affect to consumer θ. The publisher knows the age, gender, address of the consumer, as well as some other information related to his valuations for the different goods, but is not able to compute the score, because it lacks some information about the firm. Still, if the publisher reveals these characteristics to advertisers, they are able to compute the score. If the publisher decides to reveal the information, we say that it follows a *disclosure* policy. If not, we say that it follows a *privacy* policy.

Anytime a consumer visits the website, the publisher runs a second price auction in order to determine which firm will appear on the consumers' screen. For simplicity, we assume that the publisher cannot set a reserve price for the auction.

Display advertising is used both for brand building and immediate selling (See Autorité de la concurrence (2010)). Our model encompasses both dimensions. A sale can be interpreted either as a direct purchase by the targeted consumer or a purchase that happens latter on.

The timing of the game is the following:

1. The publisher commits to a policy $\sigma \in \{\mathcal{D}, \mathcal{P}\}$, where $\mathcal{D}$ stands for *Disclosure* and $\mathcal{P}$ for *Privacy*.

2. Firms choose independently and simultaneously their prices p_i.

3. Under *Disclosure*, each firm i learns θ_i. Under *Privacy*, firms do not learn θ_i.

4. Under *Disclosure*, firms can submit bids which depend on the realization of θ_i: $b_i^{\mathcal{P}}(\theta_i, p_i)$. Under *Privacy*, they submit a single bid $b_i^{\mathcal{P}}(p_i)$. The auction is a second price auction with no reserve price.

5. The consumer is matched the winning firm, say firm j. Total welfare, consumer's surplus and firm j's profit are given by $W(p_j, \theta_j)$, $S(p_j, \theta_j)$, and $\pi_j(p_j, \theta_j)$. The platform's revenue is given by the highest losing bid.

In the auction we only consider equilibria in undominated strategies, i.e in which firms bid truthfully.

3 Equilibrium under privacy and disclosure - the general case

The case of privacy Suppose that the platform chooses not to let firms learn anything. Let $P \equiv (p_1, ..., p_n)$ be the vector of prices, and P_{-i} be the vector of prices of firms other than i.

If it sets a price p_i, firm i's profit is

$$E[\pi_i^{\mathcal{P}}(p_i, P_{-i})] = \max\{\int_0^{\overline{\theta}} \pi(p_i, \theta_i) f(\theta_i) d\theta_i - T_i(P_{-i}), 0\}$$

where $T_i(P_{-i}) = \max_{j\in N-i} \int_0^{\overline{\theta}} \pi(p_j, \theta_j) f(\theta_j) d\theta_j$ is firm i's payment if it wins the auction. Notice that this payment does not depend on the realization of θ, because firms do not learn θ before they bid.

Maximizing this profit with respect to p_i leads to the following proposition:

Proposition 1 *When the platform chooses to implement a privacy policy, a symmetric equilibrium is such that the price $p^{\mathcal{P}}$ verifies:*

$$\int_0^{\overline{\theta}} \frac{\partial \pi(p^{\mathcal{P}}, \theta_i)}{\partial p_i} f(\theta_i) d\theta_i = 0 \tag{1}$$

Given that firms cannot infer anything from the fact that they win the auction, they set a price equal to the monopoly case when they have no information about consumers.

Proposition 2 *Under privacy, the platform extracts all the profits of the industry:*

$$\Pi_0 = E[\pi_i^{\mathcal{P}}(p^{\mathcal{P}})]$$

Proof: Since firms are symmetric, they all set the same price, and thus bid the same amount for every consumer. □

The case of disclosure Now we assume that before firms make a bid (but after having chosen their price), they privately learn the consumer's type. We are looking for a symmetric equilibrium, in which firms charge a price $p_n^{\mathcal{D}}$ and bid truthfully for every realization of θ.

Since firms bid truthfully, firm i's bid is $\pi(p_i, \theta_i)$.

Suppose that all the firms other than i set a price $p_n^{\mathcal{D}}$. Let $\hat{\theta}_{-i}$ be the highest realization of θ_j for $j \in N - i$. Let j_0 be the identity of the corresponding firm. By Assumption 2, i will win the auction if it bids more than firm j_0, since j_0 outbids all the other firms. Let $\phi(\hat{\theta}_{-i}, p_i, p_n^{\mathcal{D}})$ be the smallest value of θ_i such that i wins the auction. Notice that by Assumption 2, $\phi(\hat{\theta}_{-i}, p, p) = \hat{\theta}_{-i}$ for every p.

Firm i's expected profit is therefore

$$E[\pi_i^{\mathcal{D}}(p_i, p_n^{\mathcal{D}})] = \int_{\hat{\theta}_{-i}\in[0,\overline{\theta}]} \int_{\theta_i \in [\phi(\hat{\theta}_{-i}, p_i, p_n^{\mathcal{D}}), \overline{\theta}]} \left[\pi(p_i, \theta_i) - \pi(p_n^{\mathcal{D}}, \hat{\theta}_{-i})\right] f_{n-1:n-1}(\hat{\theta}_{-i}) f(\theta_i) d\theta$$

where $f_{k:m}$ is the probability distribution function of the kth order statistic of θ_j among m.[5]

At a symmetric equilibrium, we must have $\frac{\partial E[\pi_i^{\mathcal{D}}(p_i,p_n^{\mathcal{D}})}{\partial p_i}|_{p_i=p_n^{\mathcal{D}}} = 0$

This first order condition rewrites as

$$\int_{\hat{\theta}_{-i}\in[0,\overline{\theta}]} \left\{ \int_{\theta_i\in[\hat{\theta}_{-i},\overline{\theta}]} \frac{\partial \pi_i(p_n^{\mathcal{D}},\theta_i)}{\partial p_i} f(\theta_i)d\theta_i - \frac{\partial \phi(\hat{\theta}_{-i},p_i,p_n^{\mathcal{D}})}{\partial p_i} \left(\pi(p_n^{\mathcal{D}},\hat{\theta}_{-i}) - \pi(p_n^{\mathcal{D}},\hat{\theta}_{-i})\right) \right\} f_{n-1:n-1}(\hat{\theta}_{-i})d\hat{\theta}_{-i} = 0$$

After some extra manipulations, we get :

Proposition 3 *Under disclosure, a symmetric equilibrium price is given by*

$$\int_0^{\overline{\theta}} \frac{\partial \pi(p_n^{\mathcal{D}},\theta_i)}{\partial p_i} F^{n-1}(\theta_i) f(\theta_i) d\theta_i = 0 \tag{2}$$

The difference between (1) and (2) comes from the term $F^{n-1}(\theta_i)$ in the integrand. Under privacy, winning the auction for a consumer does not bring any information about the consumer's type. Under disclosure, on the other hand, under a symmetric strategy profile, firm i wins the auction only when all the θ_j's are smaller than θ_i, which occurs with probability $F^{n-1}(\theta_i)$. As we show in the next proposition, the equilibrium price is then higher under disclosure than under privacy.

Proposition 4 *For every n, the equilibrium price under disclosure is larger than the equilibrium price under privacy: $p_n^{\mathcal{D}} \geq p^{\mathcal{P}}$.*

Proof: The proof is based on a comparison of the first order conditions (1) and (2).

Let

$$\zeta^1(p) \equiv \int_0^{\overline{\theta}} \frac{\partial \pi(p,\theta_i)}{\partial p_i} f_i(\theta_i) d\theta_i$$

From (1), we have $\zeta^1(p^{\mathcal{P}}) = 0$.

Also, since by Assumption 4 we have $\frac{\partial^2 \pi_i}{\partial p \partial \theta_i} \geq 0$, then, for any increasing function h,

$$\int_0^{\overline{\theta}} \frac{\partial \pi(p,\theta_i)}{\partial p_i} h(\theta_i) f_i(\theta_i) d\theta_i \geq 0 \tag{3}$$

Now let

$$\zeta^n(p) \equiv \int_{\{\theta_i\in[0,\overline{\theta}]\}} \frac{\partial \pi(p,\theta_i)}{\partial p_i} F^{n-1}(\theta_i) f(\theta_i) d\theta_i$$

From (2), we have $\zeta^2(p_n^{\mathcal{D}}) = 0$. Using (3) with $h \equiv F^{n-1}$, one gets $\zeta^2(p^{\mathcal{P}}) \geq 0$.

[5] $f_{m:m}$ corresponds to the highest realization, $f_{m-1:m}$ to the second highest, and so on.

Moreover, ζ^n is non increasing by concavity of the profit function, and so we obtain $p^{\mathcal{P}} \leq p_n^{\mathcal{D}}$. □

The intuition for proposition 4 is straightforward: under disclosure, conditional on winning the auction, firm i expects to face consumers with a higher θ_i than under privacy, and therefore the optimal strategy is to charge a higher price.

This effect is all the more important as the number of firms is large, as the next proposition shows.

Proposition 5 *Under disclosure, the equilibrium price increases with the number of firms.*

Proof: The proof obeys the same logic as Proposition 4

Let

$$\zeta^n(p) \equiv \int_0^{\overline{\theta}} \frac{\partial \pi(p,\theta_i)}{\partial p_i} F^{n-1}(\theta_i) f_i(\theta_i) d\theta_i$$

For every n, $p_n^{\mathcal{D}}$ is such that $\zeta^n(p_n^{\mathcal{D}}) = 0$. By choosing $h_n \equiv F^{n-1}/F^{n-2} = F$, which is increasing, we get $p_n^{\mathcal{D}} \geq p_{n-1}^{\mathcal{D}}$. □

As the number of firms grows, it becomes less likely for firm i to win the auction if the consumer has a low θ_i. Therefore firms put a bigger weight on high θ_i's, which leads them to raise their price in equilibrium.

In the limit, we have the following proposition:

Proposition 6 *As n goes to infinity, the equilibrium price under disclosure tends to $p^*(\overline{\theta})$. The profit of the platform tends to $\pi(p^*(\overline{\theta}), \overline{\theta})$.*

The proof relies on the following simple idea: when the number of firms is very large, firm i knows that it will win the auction only when θ_i is very close to $\overline{\theta}$

Proof: Let $p_n^{\mathcal{D}}$ be the price if the platform chooses to implement a disclosure policy when n firms are on the market. First, notice that since $(p_n^{\mathcal{D}})_{n\geq 1}$ is non decreasing and bounded (by $\overline{p}$), it has a limit, that we note $p_\infty^{\mathcal{D}}$.

From (2), we have

$$\int_0^{\overline{\theta}} \frac{\partial \pi_i(p_n^{\mathcal{D}},\theta_i)}{\partial p_i} F^{n-1}(\theta_i) f(\theta_i) d\theta_i = 0$$

Therefore, for any n and $\epsilon \in (0, \overline{\theta})$,

$$n\left(\int_0^{\overline{\theta}-\epsilon} \frac{\partial \pi_i(p_n^{\mathcal{D}},\theta_i)}{\partial p_i} F^{n-1}(\theta_i) f(\theta_i) d\theta_i + \int_{\overline{\theta}-\epsilon}^{\overline{\theta}} \frac{\partial \pi_i(p_n^{\mathcal{D}},\theta_i)}{\partial p_i} F^{n-1}(\theta_i) f(\theta_i) d\theta_i\right) = 0$$

Let

$$A_{n,\epsilon} \equiv n \int_0^{\overline{\theta}-\epsilon} \frac{\partial \pi_i(p_n^{\mathcal{D}},\theta_i)}{\partial p_i} F^{n-1}(\theta_i) f(\theta_i) d\theta_i$$

and

$$B_{n,\epsilon} \equiv n \int_{\overline{\theta}-\epsilon}^{\overline{\theta}} \frac{\partial \pi_i(p_n^{\mathcal{D}}, \theta_i)}{\partial p_i} F^{n-1}(\theta_i) f(\theta_i) d\theta_i$$

Since $\frac{\partial^2 \pi_i}{\partial p_i \partial \theta_i} \geq 0$, one can write

$$\frac{\partial \pi_i(p_n^{\mathcal{D}}, 0)}{\partial p_i} \int_0^{\overline{\theta}-\epsilon} nF^{n-1}(\theta_i) f(\theta_i) d\theta_i \leq A_{n,\epsilon} \leq \frac{\partial \pi_i(p_n^{\mathcal{D}}, \overline{\theta}-\epsilon)}{\partial p_i} \int_0^{\overline{\theta}-\epsilon} nF^{n-1}(\theta_i) f(\theta_i) d\theta_i$$

The integral on the left side and on the right side is equal to $F^n(\overline{\theta}-\epsilon)$, and goes to zero as n goes to infinity. Therefore $\lim_{n\to\infty} A_{n,\epsilon} = 0$

By the same argument, we can provide a lower and an upper bound on $B_{n,\epsilon}$:

$$\frac{\partial \pi_i(p_n^{\mathcal{D}}, \overline{\theta}-\epsilon)}{\partial p_i}[F^n(\overline{\theta}) - F^n(\overline{\theta}-\epsilon)] \leq B_{n,\epsilon} \leq \frac{\partial \pi_i(p_n^{\mathcal{D}}, \overline{\theta})}{\partial p_i}[F^n(\overline{\theta}) - F^n(\overline{\theta}-\epsilon)]$$

Using the fact that $F^n(\overline{\theta}) = 1$, and that $A_{n,\epsilon} + B_{n,\epsilon} = 0$, we obtain

$$A_{n,\epsilon} + \frac{\partial \pi_i(p_n^{\mathcal{D}}, \overline{\theta}-\epsilon)}{\partial p_i}[1 - F^n(\overline{\theta}-\epsilon)] \leq 0 \leq A_{n,\epsilon} + \frac{\partial \pi_i(p_n^{\mathcal{D}}, \overline{\theta})}{\partial p_i}[1 - F^n(\overline{\theta}-\epsilon)]$$

By taking n to infinity, one gets

$$\frac{\partial \pi_i(p_\infty^{\mathcal{D}}, \overline{\theta}-\epsilon)}{\partial p_i} \leq 0 \leq \frac{\partial \pi_i(p_\infty^{\mathcal{D}}, \overline{\theta})}{\partial p_i}$$

If $\epsilon \to 0$, and by continuity of the derivative of the profit, we finally get

$$\frac{\partial \pi_i(p_\infty^{\mathcal{D}}, \overline{\theta})}{\partial p_i} = 0$$

i. e $p_\infty^{\mathcal{D}} = p^*(\overline{\theta})$.

The platform's profit is

$$E[\Pi_{0,n}^{\mathcal{D}}] = \int_0^{\overline{\theta}} \pi(p_n^{\mathcal{D}}, \theta_i) f_{n-1:n}(\theta_i) d\theta_i = E[\pi(p_n^{\mathcal{D}}, \theta_{n-1:n})]$$

Notice that $\theta_{n-1:n}$ converges almost surely to $\overline{\theta}$. Then, by continuity of π_i, $\pi(p_n^{\mathcal{D}}, \theta_{n-1:n})$ converges to $\pi(p^*(\overline{\theta}), \overline{\theta})$ almost surely.

By the monotone convergence theorem, we can conclude that

$$\lim_{n\to\infty} E[\Pi_{0,n}^{\mathcal{D}}] = E[\lim_{n\to\infty} \pi(p_n^{\mathcal{D}}, \theta_{n-1:n})] = \pi(p^*(\overline{\theta}), \overline{\theta})$$

□

Remark 1 *On peut prouver que* $(p_n^{\mathcal{D}})$ *converge en montrant qu'elle est croissante et majorée (par la valuation maximale). Il faut également mettre une hypothèse sur la continuité de la dérivéee du profit.*

When n is very large, which is probably the relevant case for a platform like Facebook, only the top of the distribution matters: firms do not expect their ads to be displayed to consumers for whom they are not the best match. This has implications for the paltform's optimal policy, as shown below.

Proposition 7 *When n goes to infinity, the platform's optimal policy is disclosure.*

Proof: : Under privacy, the paltform's profit is $E[\pi(p^{\mathcal{P}},\theta)]$, whereas under disclosure it tends to $\pi(p^*(\overline{\theta}),\overline{\theta})$. By Assumptions 2 and 3,

$$E[\pi(p^{\mathcal{P}},\theta)] \leq \pi(p^{\mathcal{P}},\overline{\theta}) \leq \pi(p^*(\overline{\theta}),\overline{\theta})$$

□

The trade-offs From a positive point of view, one would like to know under which conditions the platform is likely to adopt a disclosure policy. We see that under privacy, the platform extracts all the industry profits, while this is not the case under disclosure. Indeed, under disclosure, the platform's profit is $E[\pi(p_n^{\mathcal{D}},\theta_{n-1:n})]$, where $\theta_{n-1:n}$ is the random variable equal to the second highest realization of the θ_i's, whereas the industry profit is $E[\pi(p_n^{\mathcal{D}},\theta_{n:n})]$. The platform has to leave an information rent to the winner of the auction. On the other hand, even if the platform cannot extract all the profit, it is possible that the share it extracts under disclosure is higher than the whole profit under privacy.

Although the previous analysis provides some rather general insights regarding the different players' trade-offs, it does not allow us to characterize when the platform's interests are in conflict with consumers' or with social welfare maximization. In order to tackle these questions, we further specify the model by taking a simple framework that is rich enough to allow for a meaningful discussion.

4 A binary model - the case with an infinite number of firms

Suppose that for every $i \in \{1,...,n\}$, θ_i is distributed according to the cdf F over a support $[\underline{\theta},\overline{\theta}]$, and that $E[\theta]=m$. Let v_i be the consumer's valuation for product i. We assume that $v_i \in \{v_L, v_H\}$, with $\theta_i = Pr[v_i = v_H]$.

The case of privacy If the platform chooses not to disclose information to the firms, the equilibrium of the subgame is as follows:

Proposition 8 *If $mv_H \leq v_L$, the equilibrium price is $p^{\mathcal{P}} = v_L$. The platform's profit is $\Pi^{\mathcal{P}} = v_L$. Consumers' surplus is $S^{\mathcal{P}} = m(v_H - v_L)$, and social welfare is $W^{\mathcal{P}} = (1-m)v_L + mv_H$.*

If $mv_H > v_L$, the equilibrium price is $p^{\mathcal{P}} = v_H$. The platform's profit is $\Pi^{\mathcal{P}} = mv_H$. Consumers' surplus is $S^{\mathcal{P}} = 0$, and social welfare is $W^{\mathcal{P}} = mv_H$.

Proof: Under privacy, a firm expects to be match with a consumer who has a probability m of having a high willingness to pay. If $mv_H \leq v_L$, firms prefer to serve every one rather than charge a high price and serve only the v_H consumers. If $mv_H > v_L$, the opposite is true. The expressions of the platform's profit, consumers' surplus and social welfare are straightforward. □

The case of disclosure To begin with, let's assume that $n = \infty$. This allows us to have cleaner expressions and to understand some important effects. If the platform chooses to disclose information, the equilibrium of the subgame is as follows:

Proposition 9 *If $\overline{\theta} v_H \leq v_L$, the equilibrium price is $p_n^{\mathcal{D}} = v_L$. The platform's profit is $\Pi^{\mathcal{D}} = v_L$. Consumers' surplus is $S^{\mathcal{D}} = m(v_H - v_L)$, and social welfare is $W^{\mathcal{D}} = v_L + m(v_H - v_L)$.*

If $\overline{\theta} v_H > v_L$, the equilibrium price is $p_n^{\mathcal{D}} = v_H$. The platform's profit is $\Pi^{\mathcal{D}} = \overline{\theta} v_H$. Consumers' surplus is $S^{\mathcal{D}} = 0$, and social welfare is $W^{\mathcal{D}} = \overline{\theta} v_H$.

Under disclosure, if $\overline{\theta} v_H \leq v_L$, firms play a strategy that does not take advantage of the information about consumers, since the profit when a firm plays v_L is v_L, irrespective of the consumers' type. (See below for a comment on this feature of the model).

Figure 4 summarizes the optimal policy from the point of social welfare. In order to get an intuition start from a point in the lower-left area. In this area even under disclosure, firms prefer to charge v_L. This situation is clearly efficient, because the quantity traded is maximal (probability 1) and the average valuation is higher under disclosure. When $\overline{\theta}$ increases moderately, firms find it optimal to charge a high price under disclosure. This induces a distorsion in the quantity traded, which is now $\overline{\theta}$ instead of 1. Not only are consumers left with no surplus, but this distorsion makes disclosure socially inferior to privacy. As $\overline{\theta}$ further increases, a positive effect offsets the distorsion with respect to the quantity traded, namely the fact that the average valuation is much higher under disclosure than under privacy. This effect restores the optimality of disclosure. The role of v_H/v_L is analogous to the role of $\overline{\theta}$, the difference being that, when v_H/v_L is large, firms find it optimal to charge v_H even under privacy, which makes disclosure even more preferable to privacy.

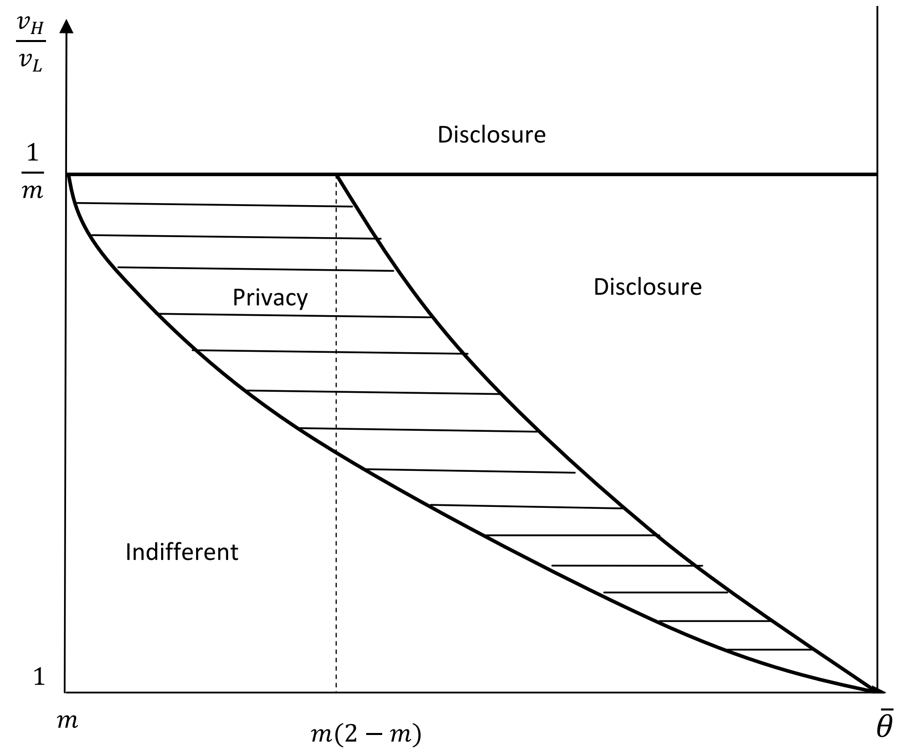

Figure 3: Socially optimal policy.

In this simple example, we see that disclosure is not always optimal from a welfare standpoint, even when the number of firms goes to infinity. There is a range of parameter values such that the distorsion that results from higher prices outweights the social benefits of better matches.

Of course, the example has several limitations. On of them is that consumers are never strictly better-off under disclosure than under privacy, because disclosure produces better matches only when firms charge a high price. Indeed, if firms charge a price equal to v_L, their expected profit is v_L irrespective of consumers' types.

One way of amending the model so as to allow for consumers to strictly benefit from disclosure would be to introduce more heterogeneity in consumers' preferences. For instance, suppose that a consumer's willingness to pay for product i belongs to $\{v_L, v_H\}$ with probability $a(\theta_i)$, and is zero otherwise. Conditionnal on this event, the probability thet $v = v_H$ is $b(\theta_i)$, with $a(.)$ and $b(.)$ non decreasing functions of θ_i. Let $\overline{a} = a(\overline{\theta})$ and $\overline{b} = b(\overline{\theta})$. Then there exists a range of parameters such that it is an equilibrium for firms to charge v_L (if $\overline{b}v_H < v_L$) . The difference with the binary model is that firms will still condition their bid on θ, so that the winning firm bids $\overline{a}v_L$. In this case, consumers strictly prefer disclosure, because it allows better matches without inducing any distorsion of trade. [6]

5 Related Literature

Our paper is related to three strands of the economics literature, namely, privacy, information revelation in auctions, and information manipulation by intermediaries.

The issue of privacy has been investigated by many recent papers both in marketing and economics through the study of targeted advertising and targeted pricing.

Targeted advertising has received increased attention in recent years. Esteban, Gil, and Hernandez (2001) show that in a monopoly framework, firms' ability to target consumers reduces both consumers' and total surplus. Roy (2000) or Iyer, Soberman, and Villas-Boas (2005) show how targeted advertising may generate market segmentation in a duopoly, respectively with homogenous and heterogenous products. In this paper we will not focus on market segmentation stemming from advertising but rather on its impact on the match improvement between the advertised good and the targeted consumer's taste. Other recent works on targeted advertising include Van Zandt (2004) who suggest ways to avoid information overload, Johnson (2008), who examines ad avoidance behavior, or Bergemann and Bonatti (2010), who study competition between medias with different targeting technologies.

[6]We are currently trying to develop a more complete example, using numerical simulations, to better capture the different effects at play.

An interesting result of the literature on targeted pricing is that in a competitive framework, firms may prefer not to have too much information about consumers, in order to avoid perfect customized pricing and hence alleviate price competition (see Thisse and Vives (1988), Chen and Zhang (2001), and Chen and Iyer (2002)). The reason is that due to imperfect information, each firm mistakenly perceives price sensitive customers as price insensitive customers, and hence reduce the rival firm's incentive to cuts prices. In our paper, we identify another reason why firms might prefer not having information about consumers to avoid customized pricing: rent extraction by an intermediary. Although each firm is an monopoly position with each consumer it gains access to, and hence, is better able to extract her surplus, the spread between the winning bid and the highest losing bid might increase with customized pricing. Consequently,the firm might eventually derive a lower net profit that under uniform pricing.

Taylor (2004) and Calzolari and Pavan (2006) focus on the sale of information by a firm who has learned something about consumers (the *upstream* firm) to a firm that interacts with them later on (the *downstream* firm). In Calzolari and Pavan (2006) an agent contracts sequentially with two principals, and the first (upstream) principal may sell whatever information he gathers about the agent to the second (downstream) principal, before the downstream principal chooses his mechanism. They focus on the incentives of the upstream principal to commit to a particular disclosure rule, and identify two effects. With the information-trade effect, the upstream principal may want to reveal information to the downstream principal since the latter is willing to pay something to acquire information. But on the other hand, when the agent knows that the upstream principal will disclose her information, and that this information will be used to reduce her rent, she needs to be given an additional incentive to reveal this information in the first place. The rent shifting effect is the following: by revealing information, the upstream principal induces the downstream principal to offer the agent a discount, and so the upstream principal can extract more rent from the agent. The authors exhibit sufficient and almost necessary conditions under which full privacy is the optimal policy from the upstream principal's point of view. Our model differ in the following respects: (i) there is no contracting between our upstream principal and the agent, only downstream contracting, and (ii) the upstream principal contracts with several downstream principals in order to allow one of them to access the agent.

In a more specified environment with consumers interacting sequentially with two firms of which goods' valuation are correlated Taylor (2004) shows that if there is no privacy and consumer are myopic, then the first firm sells the customer data to the second firm. This gives the first firm incentives to charge higher prices to better screen consumers' types and make its customer data more valuable. Firms prefer the no-privacy case, while consumers prefer the privacy case thus making privacy ambiguous for welfare.

In a recent paper, Goldfarb and Tucker (2011) estimate that privacy regulation in the EU has reduced the effectiveness of ads by 65% (measured in terms of stated purchase intent). The decline is more important on general interest websites than on specialized websites. Such empirical evidence supports the view presented here that the matching effect of disclosure is important.

In the auction literature, Ganuza (2004) and Ganuza and Penalva (2010) study the optimal amount of information disclosed by a seller prior to the auction. They identify the trade-off faced by the seller between maximizing the value from trade and minimizing the informational rent left to the winner. This trade-off is an important part of our analysis, although in our model the buyers in the auction (i.e the firms) are allowed to take an action (i.e choose a price) prior to the auction, which depends on the amount of information disclosed, and which affects the distribution of the gains from trade. In other words, their set-up corresponds to a variant of our model in which prices are chosen before the publisher commits to a disclosure policy.

This difference has important implications. One of their main results is that the seller always reveals less information than what would be optimal. In our paper this is not always true: since firms tend to increase their price when the publisher discloses information, it may be the case that disclosing information causes welfare to decrease, but that the publisher still prefers disclosure to privacy. This is true when the number of firms is large and the gains from trade (measured by v) are moderate.

Finally, our paper is also related to a recent literature which studies how intermediaries use the information in order to affect matching between firms and consumers. Hagiu and Jullien (2010), and de Cornière (2010) identify several reasons for which an intermediary may want not to implement the perfect matching. In Hagiu and Jullien (2010), if the intermediary receives a fee every time a consumer visits a firm, it has an incentive to direct consumers towards firms that they would not have visited otherwise. Another motive underlined in that paper is that the intermediary can induce firms to lower their price by manipulating the matching, which helps attracting more consumers to the platform. In a different set-up, de Cornière (2010) shows that the optimal level of accuracy of the matching solves a trade-off between consumers participation and the level of firms' profit, which can be totally captured by the intermediary. In this paper, we suggest another rationale for optimally imperfect matching, namely that implementing a good matching (by disclosing information) might be too costly in terms of informational rent left to firms.

References

Bergemann, D., and A. Bonatti (2010): "Targeting in Advertising Markets: Implications for Offline vs. Online Media," Working paper.

Calzolari, G., and A. Pavan (2006): "On the optimality of privacy in sequential contracting," *Journal of Economic Theory*, 130(1), 168–204.

Chen, Y. Narashimham, C., and Z. Zhang (2001): "Individual Marketing with Imperfect Targetability," *Marketing Science*, 20, 23–41.

Chen, Y., and G. Iyer (2002): "Consumer Addressabilty and Customized Pricing," *Marketing Science*, 21, 197–208.

de Cornière, A. (2010): "Targeted advertising with consumer search: an economic analysis of keyword advertising," Discussion paper.

Esteban, L., A. Gil, and J. M. Hernandez (2001): "Informative Advertising and Optimal Targeting in a Monopoly," *Journal of Industrial Economics*, 49(2), 161–80.

Evans, D. S. (2008): "The Economics of the Online Advertising Industry," *Review of Network Economics*, 7(3), 359–391.

——— (2009): "The Online Advertising Industry: Economics, Evolution, and Privacy," *Journal of Economic Perspectives*, 23(3), 37–60.

Ganuza, J.-J. (2004): "Ignorance Promotes Competition: An Auction Model of Endogenous Private Valuations," *RAND Journal of Economics*, 35(3), 583–598.

Ganuza, J.-J., and J. S. Penalva (2010): "Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions," *Econometrica*, 78(3), 1007–1030.

Goldfarb, A., and C. Tucker (2011): "Privacy Regulation and Online Advertising," *Management Science*, 57(1), 57–71.

Hagiu, A., and B. Jullien (2010): "Why do intermediaries divert search?," Idei working papers, Institut d'Économie Industrielle (IDEI), Toulouse.

Iyer, G., D. Soberman, and J. M. Villas-Boas (2005): "The Targeting of Advertising," *Marketing Science*, 24(3), 461–476.

Johnson, J. (2008): "Targeted Advertising and Advertising Avoidance," Working paper.

Johnson, J. P., and D. P. Myatt (2006): "On the Simple Economics of Advertising, Marketing, and Product Design," *American Economic Review*, 96(3), 756–784.

Office Of Fair Trading (2010): "Online Targeting of Advertising and Prices," *OFT 1231.*

Roy, S. (2000): "Strategic segmentation of a market," *International Journal of Industrial Organization*, 18(8), 1279–1290.

Taylor, C. (2004): "Consumer Privacy and the Market for Customer Information," *RAND Journal of Economics*, 35, 631–650.

Thisse, J., and X. Vives (1988): "On the Strategic Choice of Spatial Price Policy," *American Economic Review*, (78), 122–137.

Van Zandt, T. (2004): "Information Overload in a Network of Targeted Communication," *RAND Journal of Economics*, 35(3), 542–560.

Vives, X. (2001): *Oligopoly Pricing: Old Ideas and New Tools*, vol. 1 of *MIT Press Books*. The MIT Press.