The value of consumer data in online advertising*

Marc Bourreau† Bernard Caillaud‡ Romain De Nijs§

June 7, 2018

Abstract

In this paper we propose a model where consumer personal data have multidimensional characteristics, and are used by platforms to offer ad slots with better targeting possibilities to a market of differentiated advertisers through real-time auctions. A platform controls the amount of information about consumers that it discloses to advertisers, thereby affecting the dispersion of advertisers' valuations for the slot. We first show by way of simulations that the amount of consumer-specific information that is optimally revealed to advertisers increases with the degree of competition on the advertising market and decreases with the cost of information disclosure for a monopolistic platform, competing platforms or a welfare-maximizing platform, provided the advertising market is not highly concentrated. Second, we exhibit different properties between the welfare-maximizing situation and the imperfectly competitive market situations with respect to how the incremental value of information varies: there are decreasing social returns to consumers' data, while private returns may be increasing or decreasing locally.

*We thank an anonymous referee and the Editor, Alexandre de Cornière, for very helpful remarks and suggestions. This research has been financed by France Strategie, within the general framework of a Research Project on the "Evolution of the Value created by the Digital Economy and its Fiscal Consequences", signed between France Strategie and a consortium consisting of PSE and Telecom ParisTech.

†Telecom ParisTech; Email: `marc.bourreau@telecom-paristech.fr`.

‡Paris School of Economics - Ecole des Ponts ParisTech; Email: `caillaud@pse.ens.fr`.

§CREST - Ecole Polytechnique; E-mail: `romain.de-nijs@polytechnique.edu`.

1 Introduction

The increasing digitalization of the economy goes with an increasing concern by public authorities about the use of consumers' personal data that are collected massively by many major online players.[1] Whenever they surf on the internet, visit a website, login on an account, answer various requests about personal information, choose a product or some entertainment service and make payments, consumers leave footprints in the system that may be collected in various ways: tracking cookies, device fingerprinting, history sniffing methods, etc.

Consumers' personal data are a source of value in at least three different ways.[2] First, they enable online firms to adapt their proposals to the consumer. Online retailers or service providers can propose more personalized goods or services that better match with the consumer's characteristics and implicit tastes, and a better fit generates a larger surplus to grab. Along with more personalized services, personal data may be used to identify the consumer's willingness to pay for a given (possibly personalized) product, opening the door to personalized pricing, i.e. third-degree or even first-degree price discrimination, by online sellers. Second, consumers' personal data can be used to design personalized services that are external to the visited website: in this vein, the major use of data is through the more efficient targeting of advertising messages, i.e. through advertising messages that contain personalized elements increasing the probability of reaching the consumer, of raising his/her attention and of inducing him/her to look for the advertised product and ultimately make a purchase. Third and last, consumers' data may be sold to intermediaries – the so-called data brokers – that accumulate data from various sources, consolidate and process them so that they can be sold back to other online players, websites or ad networks, in a useful format.

The mere fact of collecting, using and selling personal information about consumers is problematic as it may constitute a major violation of consumers' privacy.[3] Offering more adequate products or more relevant ad messages, absent any impact on prices, is presumably welfare improving and bring better value to the consumers, but the increasing capacity of online players to extract value from consumers through more sophisticated price discrimination may reduce or even reverse the direct effect of personalization of services and of advertising. A precise analysis of the impact of the availability of consumers' personal data in online industries therefore calls for a more thorough study of how data can be turned into value, how a market for data works and how data is transferred or

[1] See e.g. the reports for the U.K. of OFT (2010) and (2013), for the US FTC (2014) or the French Autorité de la Concurrence and Bundeskartellamt's report (2016).

[2] On the business models of data-driven platforms, see, for example, Lambrecht et al. (2014).

[3] For a recent survey on the economics of privacy, see Acquisti et al. (2016).

exchanged among players.

This paper aims at addressing the second motive for the use of personal data, i.e. the fact that data can help ad networks or operators to improve ad targeting and therefore the efficiency of advertising campaigns. Typically, advertising slots that are consumer-specific are sold in real time through auctions and bidders are able to use some of the personal data about the consumer that is attached to the slot so as to better assess the value of the slot for them and how much they are willing to bid for it. The platform that runs the auctions and that has collected information about the specific consumer can control the information revealed to bidders so as to extract maximal value from the auction. Of course, competition among platforms to access the consumer's attention may mitigate this objective and lead the platform to modify its disclosure policy so as to fight for market shares on the advertising market.

In this setting, we investigate what is the value of information for data collecting platforms, depending on the market structure at the platform level and on the structure of the advertising market. More precisely, we analyze how the value of information and the incremental value of information, i.e. the value of an additional piece of information, vary as functions of the degree of competition in the advertising market, in three situations: when the platform is monopolistic, when there is competition between two platforms in a competitive bottleneck situation (see Armstrong, 2006), and finally when a central planner aims at maximizing social efficiency. The study of the incremental value of information relates to whether there are increasing or decreasing returns in the use of consumers' personal information. Ultimately, these issues lead to the question of the incentives of monopolistic, competing or regulated platforms to collect information about consumers.

One important aspect of our setting is that we study a model that simply formalizes what disclosure of an additional piece of personal information on consumers means for advertisers and that accounts for the fact that two differentiated advertisers may evaluate differently some revealed set of personal data. For this, we have to depart from models that rely on all or nothing disclosure that lead to prior information or perfect information ex post, and we have to enrich the view of information as not being reducible to a one-dimensional variable. So, we examine a simple model in which information on consumers is multidimensional and directly related to the number of consumer characteristics that are disclosed to advertisers. In this setting, additional pieces of information may increase or decrease advertisers' value and information revelation is related to the dispersion of bidders' interim evaluations in the advertising market. Although the model has many attractive features, its complete analytic analysis is not tractable and in this paper we use simulations to discuss the main properties of the model.

We first show that the amount of consumer-specific information that is optimally revealed to advertisers in the three situations (a monopolistic platform, two competing platforms and a welfare-maximizer) increases with the degree of competition on the advertising market, i.e. with the number of advertisers, as long as there are at least 4 advertisers. The intuition for the monopolist and the welfare-maximizer cases is the following one. Disclosing information increases the dispersion of valuations among bidders. When there are enough draws in this distribution, both the highest valuation and the second highest valuation increase with dispersion, as well as the difference between the two, as the probability of a higher value fit increases. Under competition, platforms compete to attract all single-homing advertisers. Therefore, their incentive is to disclose as much information as possible as long as they still make a non-negative profit to increase the expected value for advertisers to join them. When there are more advertisers to attract, the expected second-highest bid increases and hence platforms can incur a higher cost (under the non-negative profit condition) by providing more information to advertisers. This explains why under competition information revelation increases with the number of advertisers.

Also, when there is a cost of revealing a piece of information, we find that the amount of consumer-specific information that is optimally revealed to advertisers in the three situations decreases with the cost of information revelation. Secondly, we compare the amount of revealed information across the three situations: competing platforms fight to attract single-homing advertisers and therefore reveal more information than a monopoly platform; a monopoly platform itself reveals too little information compared to the social optimum, since it is interested in the second highest valuation and not the highest valuation; however, we find no systematic relationship between information disclosure by competing platforms and the socially efficient disclosure policy.

As for the incremental value of information about consumers, we find a contrast between the welfare maximizer situation and market situations. There exists decreasing social returns to consumers' data, while private returns may be either decreasing or increasing. Similarly, the incremental social value of consumer information increases with the degree of competition in the advertising market (in the number of advertisers in the market), while the incremental private value can either increase or decrease with it.

Our paper is related to the literature on information revelation in auctions. Our model is directly based on Ganuza (2004) and Ganuza and Penalva (2010), and most of our results are reminiscences of theirs. In particular, these articles have already shown that a monopoly auctioneer reveals less information than what would be socially desirable, and that the amount of information revealed is increasing in the degree of competition. Our main contribution with respect to these articles is to give a clear definition of the incremental value of an additional piece of information in the context of programmatic advertising thanks to a framework with an explicit formalization of consumer data as a vector of characteristics. In addition, they do not consider the case of competing auctioneers as we do in this paper. Another closely related article is De Corniere and De Nijs (2016), which endogenizes the value of additional information via the pricing strategies of bidders on a final downstream market. In the present article, we do not take into account this aspect. Last, our article is related to Board (2010) and Troncoso-Valverde (2017). Board (2010) shows that an auctioneer is always better off revealing information to all bidders in a second-price auction as long as there are at least three bidders. Troncoso-Valverde (2017) shows that when two auctioneers compete for the two bidders that are active on the market, they reveal information about their object, a result that is opposed to what is found for a monopolist auctioneer. A key difference between our article and Troncoso-Valverde (2017) is that in his setting bidders receive the signals from auctioneers before choosing which one to patronize while in our they receive it after.

The paper is organized as follows. The model is presented in Section 2, and the specific auction mechanism with multi-dimensional types is briefly analyzed in Section 3. Section 4 gathers the main results of the paper about the decision to reveal information by the platform(s). The last section concludes.

2 Model

We consider a model based on Ganuza (2004) and Ganuza and Penalva (2010), where advertising-supported digital platforms sell their ad slots to advertisers via a second-price auction. An ad slot is specific to a consumer and a platform has to decide how much information about the consumer it should reveal to bidders. We present below the different players and their strategies. The market structure is represented in Figure 1.

Consumers

There is one (representative) consumer, defined by a vector of C characteristics. The consumer's characteristics are independently and identically distributed: each characteristic takes value -1 with probability $1/2$ and 1 with probability $1/2$. A characteristic can correspond to a demographic information (young vs. old), an interest (likes fishing vs. does not like fishing), etc.

Digital platforms

Digital platforms act as intermediaries between the representative consumer and advertisers. They derive revenue solely from advertising. We consider two possible market structures: either the market is dominated by a monopoly platform or there are $M \geq 2$ competing platforms. If the platform is a monopoly, we assume full consumer participation. In the oligopoly case, we assume that the representative consumer multi-homes and that all advertisers single-home in the same platform. On the consumer's side, the idea is that the consumer can access and use all platforms free of charge,[4] and because of some differentiation between the platforms, she decides to multi-home.[5] On the advertisers' side, we consider that only the first impression of an advertisement is valuable, which implies that advertisers post an ad on one platform only, hence single-home.

A platform sells an ad slot displayed to the representative consumer to n potential advertisers via a second-price auction, where n is assumed exogenous. Each platform has perfect information about the consumer (*i.e.* it knows all C characteristics) and has to decide how much of this information to reveal to advertisers, that is, the number of characteristics c it wishes to disclose, with $0 \leq c \leq C$. In our model, the number of consumers' characteristics revealed to advertisers corresponds to the precision of the signal in Ganuza (2004) and Ganuza and Penalva (2010). Note that once we know the number of characteristics that the platform optimally reveals to advertisers, we also have information on how much data about the consumer the platform has an incentive to collect.

We assume furthermore that the platform incurs a cost δ for each characteristic revealed to the advertisers. Hence, if the platform reveals c characteristics, its cost of information revelation is δc.[6] For example, we can consider that the consumers dislike that their personal information be disclosed to advertisers and that the platform has to compensate them to ensure their participation, either through a lower price or a higher quality of service, both of which are costly.

Advertisers

There are $n \geq 2$ advertisers. Similar to the representative consumer, each advertiser is defined by a vector of C independently and identically distributed characteristics, where

[4] For example, the platforms may offer a basic service for free that attracts consumers, such as news, weather information, email, etc.

[5] One could argue that there are attention costs for the consumer of joining two platforms instead of one. We assume that the benefits from multi-homing in terms of increased variety outweigh such costs.

[6] The information revelation cost is incurred only when the consumer characteristics are revealed. If a platform announces that it will reveal some consumer characteristics to advertisers, but the advertisers join the rival platform, the information cost is not incurred.

each characteristic takes either value -1 or 1, with equal probability. The vector of characteristics of an advertiser is private information to this advertiser and in particular it is unknown to the platform.

The value of the match between a given advertiser and the representative consumer is additive across characteristics and is simply defined as the scalar product of their vectors of characteristics.

It is worth discussing at this stage the main effects at play in the model. If the platform does not reveal any information about the consumer, all advertisers have the same expected valuation for the ad slot. In contrast, by revealing information about the consumer, the platform creates dispersion among advertisers in their valuation for the slot. This dispersion increases the willingness to pay of some advertisers, which increases the platform's expected revenues and the gain from trade between the consumer and the advertiser. However, increased dispersion also creates an informational rent for advertisers. Therefore, the platform faces a trade-off when it decides on the amount of information to disclose between increasing willingness to pay and reducing the informational rent of advertisers.

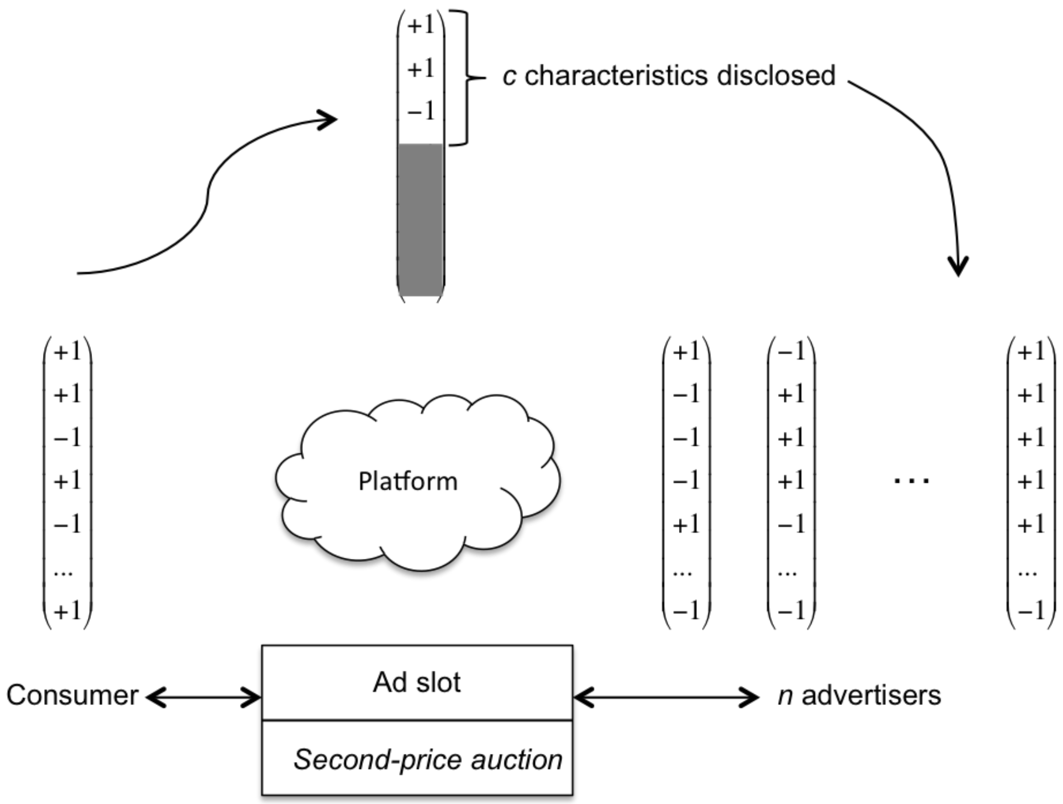

Figure 1: The market structure.

3 Second-price auction for ad slots

We start by calculating the relevant statistics of the second-price auction for the ad slot, for a given number $c \leq C$ of consumer characteristics revealed by the platform to advertisers.

Let $\tilde{V}(c)$ denote the random variable that corresponds to the value for an advertiser of displaying an ad to the consumer, given that c characteristics have been disclosed. The value of a match for a given characteristic is 1, and the value of a non-match is -1. Note that the expected value for the characteristics that are not revealed is equal to 0 (since $1/2 \times (-1) + 1/2 \times 1 = 0$). The random variable corresponding to the value of the ad slot for an advertiser, $\tilde{V}(c)$, therefore takes value in $\{-c, -c+2, \cdots, c\}$. More precisely, if $j \leq c$ characteristics match between the advertiser and the consumer, the random variable $\tilde{V}(c)$ takes value $x_j(c) \equiv (1) \times j + (-1) \times (c-j) = -c + 2j$ for $j = 0, 1, ...c$.

Let $p_j(c)$ denote the probability that the value of an ad for the consumer takes value $x_j(c)$ when c characteristics have been disclosed. This probability is equal to the number of ways to choose j characteristics that match among a total of c characteristics, divided by the number of different vectors with c characteristics. Therefore, we have:

$$p_j(c) = \frac{1}{2^c}\binom{c}{j}.$$

The random variable $\tilde{V}(c)$ therefore takes value $x_j(c) = -c+2j$ with probability $p_j(c)$, for $j = 0, 1, ...c$. The mean is null and the dispersion of this random variable increases with the number of characteristics c that are revealed.

We model the auction for the ad slot as a second-price auction: the highest bidder wins the auction, and pays the second highest bid for the ad slot.

We now derive the density functions for the largest and second-largest bidder's value for an ad to the consumer. Define $P_j(c)$ as the discrete cumulative distribution function of $\tilde{V}(c)$, with $P_0(c) = 0$ and $P_j(c) = p_1(c) + p_2(c) + ... + p_j(c)$. Omitting the dependency upon c, the discrete probability distribution of the second-largest bidder is then given by:[7]

$$Prob\{V_{(n-1)} = x_j\} = \sum_{k=n-1}^{n} \binom{n}{k} \left[P_j^k(1-P_j)^{n-k} - P_{j-1}^k(1-P_{j-1})^{n-k}\right],$$

that is,

$$Prob\{V_{(n-1)} = x_j\} = n\left[P_j^{n-1}(1-P_j) - P_{j-1}^{n-1}(1-P_{j-1})\right] + P_j^n - P_{j-1}^n. \tag{1}$$

[7] On order statistics with discrete distributions, see, for example, http://www.math.ntu.edu.tw/~hchen/teaching/StatInference/notes/lecture37.pdf.

Similarly, omitting c, the discrete probability distribution for the largest bid is given by:

$$Prob\{V_{(n)} = x_j\} = P_j^n - P_{j-1}^n. \tag{2}$$

4 Information revelation by the platforms

In this section, we study the information revelation decision by the platforms. To begin with, we compute the surplus that the different parties derive from the ad slot. Then, we study how the market structure of the advertising market affects information disclosure, comparing the environment with a monopoly platform and the environment with competing platforms. We also analyze the value of an incremental piece of information disclosed to the advertisers, and its determinants. Finally, we discuss the impact of the cost of information on disclosure.

4.1 Gains from trade

The surplus that a monopoly platform expects to derive from the sale of its ad slot is equal to the expected price of the slot, which we denote by v_m, minus the information cost, δc. From the analysis of the second-price auction in the previous section, the expected price of the ad slot is equal to the expected value of the second-highest bid, that is,

$$v_m(c) = \sum_{j=0}^{c} x_j(c) Prob\{V_{(n-1)}(c) = x_j(c)\},$$

where $Prob\{V_{(n-1)}(c) = x_j(c)\}$ is defined in (1). The expected profit of the platform is then $\pi(c) = v_m(c) - \delta c$.

Notice that the expected profit of the platform, $\pi(c)$, depends on the number of consumer characteristics disclosed to the advertisers. In particular, if it reveals no consumer information to the advertisers (*i.e.* $c = 0$), the platform makes zero profits ($\pi(0) = v_m(0) = 0$ as it can be easily checked).

The expected net surplus of the advertiser that wins the auction is equal to the difference between its value for the ad slot and the price it pays for it. Therefore, it is equal to the difference between the highest expected bid and the second-highest expected bid. Using (2), the expected value of the highest bid is equal to

$$v_1(c) = \sum_{j=0}^{c} x_j(c) Prob\{V_{(n)}(c) = x_j(c)\},$$

and the expected net surplus for the advertiser that wins the auction is then $v_c(c) =$

$v_1(c) - v_m(c)$. If the platform reveals no information about the consumer (*i.e.* $c = 0$), advertisers all have the same expected value from the match, *i.e.* 0, and therefore, we have $v_c(0) = 0$.

Finally, for the social planner, the expected value from the sale of the ad slot is equal to the expected value of the highest bid, $v_1(c)$, minus the information cost, δc. That is, $w(c) = v_1(c) - \delta c$.

In what follows, we interpret v_m as the *private* value of information and v_1 as the *social* value of information.

The monopoly platform chooses a number c of consumer characteristics to disclose to the advertisers in order to maximize its expected advertising profit, $\pi(c) = v_m(c) - \delta c$.

Under platform competition, the $M \geq 2$ platforms, compete to attract single-homing advertisers. Each platform i announces simultaneously the number of consumer characteristics, c_i, it will disclose to advertisers, with $c_i \in \{0, 1, \cdots, C\}$. Then, the advertisers decide which platform to join. In order to simplify the characterization of the equilibrium, we make the restrictive assumption that advertisers coordinate their decision and all join one platform, for example because they are coordinated by an ad network. More precisely, all advertisers join one of the platforms that offers them the largest expected net surplus. This means that if there are $0 \leq K \leq M$ platforms which offer the highest expected profit for advertisers in the market, all advertisers join one of these platforms with probability $1/K$. In such a case, these K platforms make an expected profit $\pi(c) = (1/K)(v_m(c) - \delta c)$. The other platforms make zero profit.

In the following, some of the results we propose can be proved analytically and without resorting to the specific combinatorial model we presented.[8] But this is not the case for all of our results. The advantage of our combinatorial model is that, for any set of exogenous parameters (n, c, δ), it is possible to compute all (expected) payoffs numerically and therefore to determine the amount of information disclosed to advertisers in our different scenarios. So, in the rest of the paper, we revert to numerical simulations to derive all our results and we indicate, when possible, the analytical proofs of the more general results.

More precisely, we checked the validity of each of our results below (even those that we prove analytically) by running numerical simulations with the following assumptions. First, since the expected value of undisclosed characteristics is equal to 0, the total number of consumer characteristics can be set to $C = 10$ without loss of generality. The number of characteristics disclosed to advertisers, c, then belongs to $\{0, 1, ..., 10\}$. We set the number of competing platforms to $M = 2$. The objective functions of the private platforms and of the welfare-maximizing platform depend only on the number of characteristics disclosed, c, the number of advertisers, n, and the cost of information, δ.

[8] We thank a reviewer of this journal for pointing this out.

For our simulations, we consider cases in which the number of advertisers n assumes some value in $\{2, 3, ..., 20\}$, and that the information cost parameter δ assumes some value in $\{0.1, 0.2, ..., 0.9\}$.[9]

The three following results are useful for the rest of the analysis.

Result 1 : *The social value of information, v_1, is increasing in c for all $n \geq 2$.*

Proof. Consider the random variable that characterizes the value of an ad for one advertiser as a function of a list of c characteristics being disclosed. Disclosing one more element amounts to adding up a white noise to this random variable, since it replaces a zero-value for some characteristic by +1 or -1 with probability 1/2 each. So, $\tilde{V}(c+1)$ is a mean-preserving spread of $\tilde{V}(c)$.[10] The vector of valuations for the n advertisers when $c+1$ characteristics are disclosed is therefore a mean-preserving spread of the same vector of valuations when only c characteristics are disclosed.

Taking the maximum of these n valuations is a convex transformation. Taking the expectation of a convex transformation of a mean-preserving random variable leads to an increase in value; hence, $v_1(c) \leq v_1(c+1)$. Finally, the inequality is strict for $n \geq 2$, since the advertisers' valuations are non-degenerate and i.i.d. ■

Intuitively, the larger the number of consumer characteristics disclosed to the advertisers, the better the match between the consumer and the winning advertiser, and therefore the higher the social value of information.

Result 2 : *The private value of information, v_m, decreases with the amount of information disclosed if $n = 2$, is constant with respect to c and equal to 0 if $n = 3$, and is increasing in c if $n \geq 4$. Moreover, it increases with n, the number of advertisers, if c is strictly positive.*

Proof. If $n = 2$, the second-highest random valuation is in fact the minimum of two random variables, a concave transformation. Using the same argument as in the proof of Result 1, it follows that $v_m(c) \geq v_m(c+1)$.

If $n = 3$, note that the random vector of advertisers' valuations takes value (V_1, V_2, V_3) and $(-V_1, -V_2, -V_3)$ with equal probabilities. In one case, the second highest valuation is equal to (-1) times the second highest valuation in the other case. Therefore, the expected value of the second-highest valuation has to be equal to 0.

If $n \geq 4$, it is interesting to note that taking the second-highest valuation is neither a convex nor a concave transformation of the random vector of advertisers' valuations,

[9]Note that the maximum value of a consumer characteristic is 1. So, we need to assume that $\delta < 1$, as otherwise a platform would never disclose any information to advertisers.

[10]More precisely, for each list; but given the symmetry among characteristics, we omit this detail.

so that no general argument can be made to compare the expected value for c and $c+1$ characteristics.

■

From the point of view of the monopoly platform, if there are only two advertisers, revealing some information will decrease the second-highest bid below the mean valuation, i.e., zero. Therefore, the platform is better off revealing no information. If $n = 3$, the expected value for the second-highest bidder is just equal to the mean valuation (*i.e.* 0). Finally, if $n \geq 4$, disclosing more information increases the expected value for the second-highest bidder, and hence the advertising revenue.

From Results 1 and 2, $v_1(c)$ increases in c, as well as $v_m(c)$ (if $n \geq 4$), and therefore the relation between the winner advertiser's expected surplus, $v_1(c) - v_m(c)$, and the amount of information disclosed, c, is a priori ambiguous. However, we have the following result.

Result 3 : *The expected surplus of the winning advertiser,* $v_1(c) - v_m(c)$*, increases with the amount of information disclosed.*

A consequence of Result 3 is that under platform competition, advertisers join one of the platforms that reveal the highest amount of information.

When $n \geq 4$, the game of competition in information disclosure levels has then a symmetric equilibrium, where all platforms set the highest disclosure level for which their profit is non-negative, that is, $c^\star = \min\{C, \bar{c}\}$, where $\bar{c}$ is the highest amount of information such that $v_m(c) - \delta c \geq 0$. Indeed, no platform has an incentive to disclose less information (*i.e.* $c' < c^\star$), since advertisers would then leave the platform and it would obtain zero profit. Providing more information is not a profitable deviation either, since by the definition of $c^\star$ the deviating firm would obtain a negative profit (or this deviation is not possible, if we have the corner solution, $c^\star = C$).

Is it the unique equilibrium? Not necessarily. First, note that there is no asymmetric equilibrium, as the platform with the lowest disclosure level would then deviate by providing at least the same amount of information as its rivals. There can be other symmetric equilibria, though, with lower disclosure levels.

For example, consider the possibility of an equilibrium at $\hat{c} \equiv c^\star - 1$. There is no profitable deviation to a lower c, as explained above. However, a platform may have an incentive to deviate from $\hat{c}$ to $\hat{c} + 1 = c^\star$, if:

$$v_m(\hat{c}+1) - \delta(\hat{c}+1) > \frac{1}{K}\left(v_m(\hat{c}) - \delta\hat{c}\right). \tag{3}$$

If (3) holds, there is no equilibrium at $\hat{c}$. This is true in particular if $\delta = 0$, and we expect that this is true more generally if the cost of information revelation δ is low enough. This is also true when K goes to infinity, which means that when there are more competition between platforms, our equilibrium is more likely to be the unique one. This also means that when there are more competition between platforms, they tend to reveal more information because possible equilibria with lower levels of information provision than $c^\star$ become less likely. If (3) does not hold, there is an additional equilibrium at $\hat{c}$. By the same reasoning, there may be other equilibria with lower levels of information disclosure.

In our simulations, we find that in many cases (but not all), $c^\star$ is the unique equilibrium of the game of information disclosure, in particular when the cost of information disclosure is low. For our comparative statics described below, we therefore select $c^\star$ as the equilibrium under platform competition.

4.2 Information revelation and the advertising market

We now study how the market structure of the advertising market (*i.e.* the number of advertisers, n) affects information revelation.

We begin by studying how the socially-optimal amount of information varies with the number of advertisers. We obtain the following result.[11]

Result 4 *: The socially-optimal amount of information is increasing in the number of advertisers.*

This result corresponds to Proposition 1 in Ganuza (2004) or Corollary 2 in Ganuza and Penalva (2010). The social planner cares about the value of the ad slot for the winning advertiser, $v_1(c)$, and the cost of information disclosure, δc. When the number of advertisers increases, the expected value of the match increases, since there is a higher probability of a good match. This increases the social incentives to avoid misallocating the add slot by providing more information to advertisers.

We now turn to the information revelation decision of a monopoly platform. We obtain the following result.

Result 5 *: A monopoly platform (i) does not reveal any consumer information if $n = 2$ or $n = 3$; (ii) for $n \geq 4$, it reveals an amount of information increasing in the number of advertisers.*

[11]In order to provide an analytical proof for the results pertaining to how the amount of information optimally disclosed in any situation varies with the number of advertisers, one should be able to show that the expected value of the highest (for v_1) or the second highest (for v_m) valuation among the n i.i.d. random valuations for advertisers is supermodular in (n, c). We have not been able to show this analytically for the general cases, i.e., when $n \geq 4$.

Proof. Point (i) is a direct consequence of Result 2 above: the auction expected revenue does not increase in c so that the monopolist should not disclose any characteristics. ■

This result corresponds to Proposition 3 in Ganuza (2004) or Corollary 3 in Ganuza and Penalva (2010). Though the objective function of the monopoly platform is different from that of the social planner - the former maximizes the second-highest bid and the latter the highest bid, all net of the information costs - the intuition is similar to the one provided for the previous result.

Finally, we analyze the amount of information revealed under platform competition. We obtain the following result.

Result 6 *: Under platform competition, (i) the platforms reveal no consumer information if $n = 2$ or $n = 3$; (ii) for $n \geq 4$, the amount of information revealed in equilibrium is increasing in the number of advertisers.*

Proof. Again here, point (i) is a direct consequence of Result 2 above: since the auction expected revenue is non-positive for $n = 2$ or $n = 3$, no platform will disclose any information about characteristics.

When $n \geq 4$, $v_m(c)$ is increasing in n. Therefore $\bar{c}$ is non-decreasing in n. Since competition leads to equilibrium number of disclosed characteristics equal to $c^{\star} = \min\{C, \bar{c}\}$, (ii) follows.

■

Under competition, platforms compete to attract all single-homing advertisers. Therefore, their incentive is to disclose as much information as possible as long as they still make a non-negative profit to increase the expected value for advertisers to join them. When there are more advertisers to attract, the expect second-highest bid increases and hence platforms can incur a higher cost (under the non-negative profit condition) by providing more information to advertisers. This explains why under competition information revelation increases with the number of advertisers.

Finally, we compare the information disclosure with a monopoly platform and with competing platforms to the social optimum.

Result 7 *: (i) the amount of information revealed by competing platforms is higher than with a monopoly platform; (ii) a monopoly platform reveals less information than what is socially optimal; (iii) competing platforms reveal less or more information than what is socially optimal.*

Proof. The number of characteristics revealed under monopoly necessarily yields a non-negative profit, hence is smaller than or equal to $\bar{c}$. This proves point (i). ■

Under platform competition, the platforms compete in information disclosure to attract single-homing advertisers. It is therefore natural to find that information disclosure is higher with competing platforms compared to a monopoly platform.

The second result is due to the fact that the monopoly platform maximizes the value for the second-highest bidder, which is lower than the value for the highest bidder. Since the marginal cost of information disclosure is constant and the same for the social planner and the monopoly platform, the latter ends up disclosing less information than the former.

Finally, the third result shows that platform competition can reduce inefficiency in some cases, but can also generate another type of inefficiency when the competing platforms disclose too much information compared to the social optimum.

Figure 2 offers a numerical example of Results 4-7. It first shows that the amount of information disclosed by a monopoly platform, competing platforms, and the social planner, are all increasing with the number of advertisers in the market (Results 4-6). Second, it shows that the amount of information disclosed is higher with competing platforms than with a monopoly platform, and that the amount of information disclosed under competition can be lower or higher than socially-optimal (Result 7).

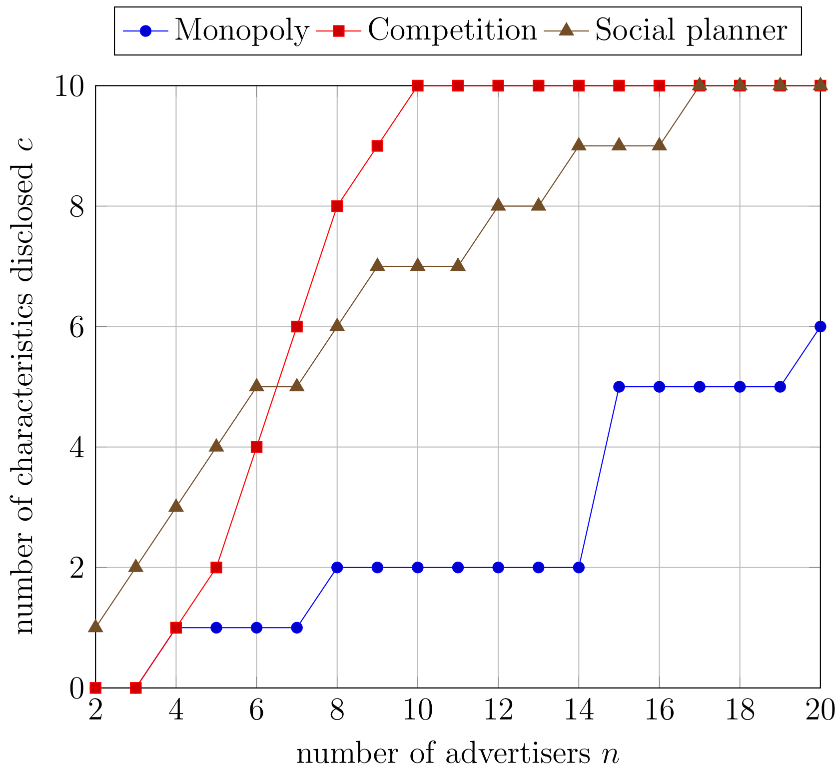

Figure 2: Number of characteristics disclosed as a function of the number of advertisers, for $\delta = 0.3$.

4.3 Incremental value of information

Now, we study how the incremental value of an additional piece of information, i.e., $v(c+1) - v(c)$, varies with (i) the amount of information already disclosed (i.e., c), and (ii) the number of advertisers in the market (i.e., n).

The first question asks whether we should expect increasing or decreasing returns in the value of additional data and the second question whether the market structure in the advertising market can affect the incremental value of data.

We have the following results.[12]

Result 8 : *The incremental social value of consumer information is decreasing in the number of characteristics disclosed.*

Figure 3 provides an illustration of this result. The figure shows how the incremental social value of information, $v_1(c+1) - v_1(c)$, varies with c for different numbers of advertisers (5, 15 and 20). We observe that this incremental value is always decreasing with the number of characteristics disclosed, c. This result therefore suggests decreasing social returns to consumer data.

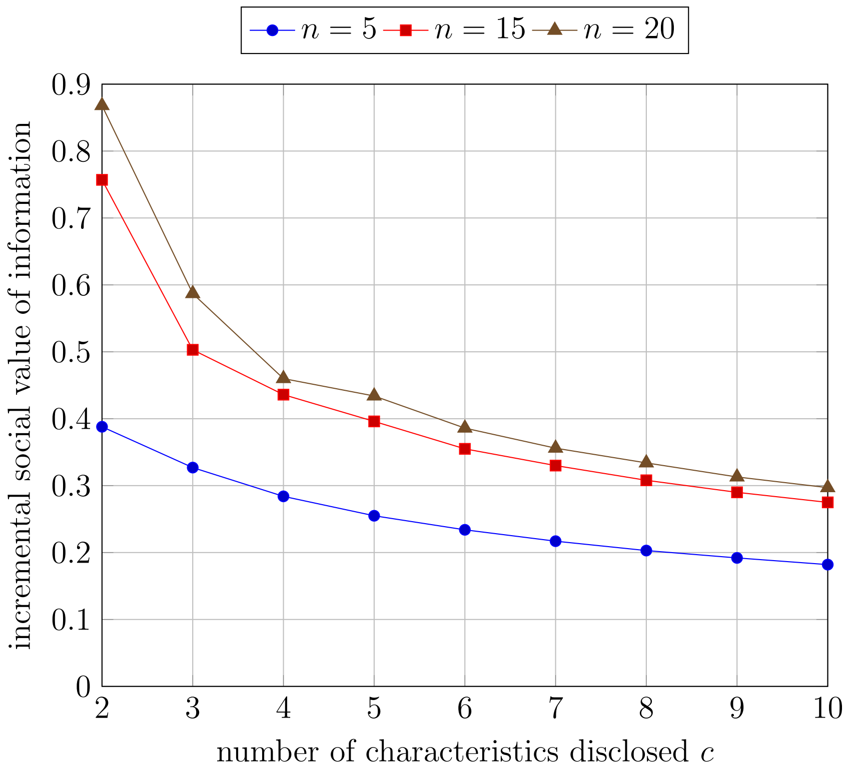

Figure 3: Incremental social value as a function of the number of characteristics disclosed.

[12]We could not come up with analytical proofs for the results in this sub-section, but we conjecture that such proof cannot be easily obtained as these results pertain to the concavity or convexity, and not to monotonicity properties, of the value of information disclosure.

Result 9 *: The incremental private value of consumer information varies non-monotonically with the number of characteristics disclosed.*

Figure 4 shows how the incremental private value of information, $v_m(c+1) - v_m(c)$, varies with c for the same numbers of advertisers as in Figure 3 (i.e., 5, 15 and 20). As we can see, the incremental private value can either increase or decrease in c. This result therefore suggests that there is no clear increasing or decreasing private returns to consumer data. The relation between the incremental value of data and the amount of data already disclosed is more complex.

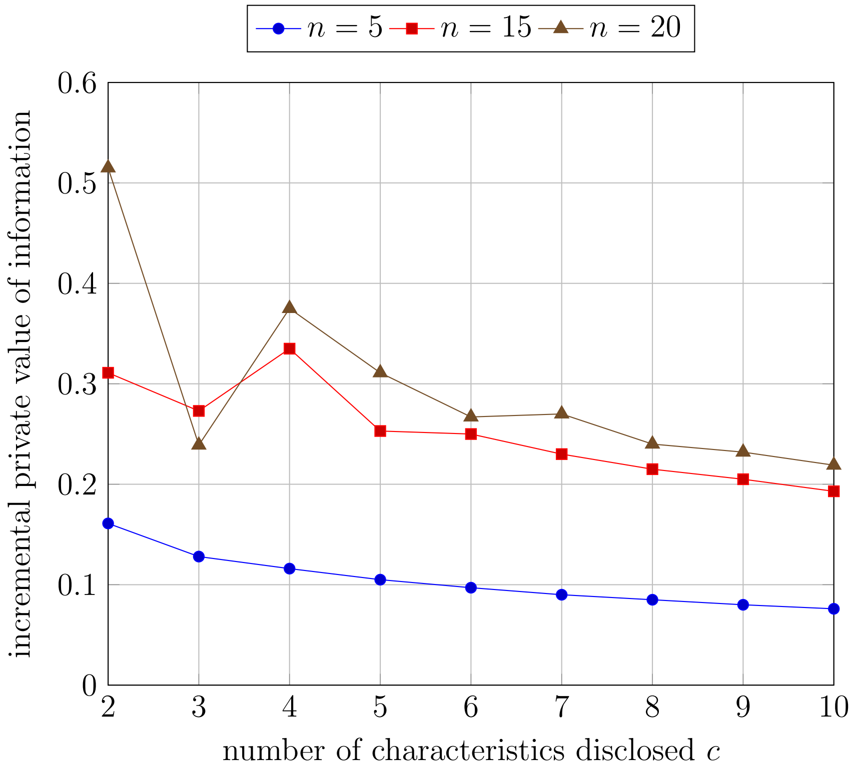

Figure 4: Incremental private value as a function of the number of characteristics disclosed.

Finally, we analyze how the incremental value varies with the number of advertisers. We have the following result.

Result 10 *: The incremental social value of consumer information increases with the number of advertisers in the market. The incremental private value can either increase or decrease with it.*

Figure 5 provides an example of the private and social incremental value of consumer information as a function of the number of advertisers, n, for $c = 3$. We observe that the incremental social is increasing for all n, whereas the incremental private value is first increasing then decreasing in n.

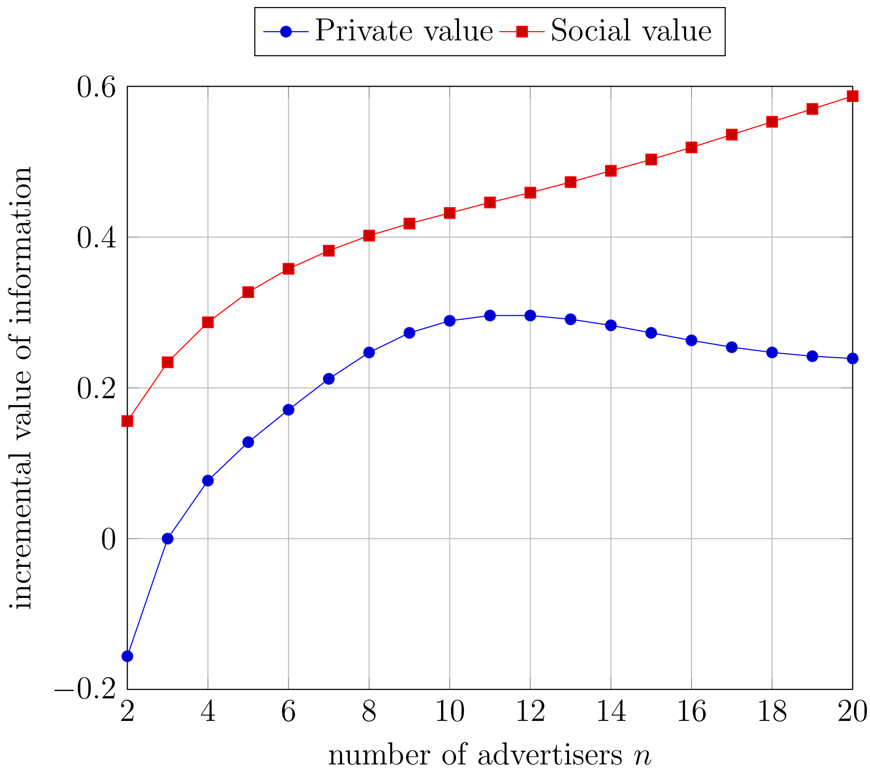

Figure 5: Incremental value as a function of the number of advertisers, with $c = 3$.

4.4 Cost of information

Finally, we study the impact of the information cost δ on the amount of information disclosed by the social planner, a monopoly platform, and competing platforms.

Result 11 : *The amount of information disclosed by the social planner, the monopoly platform, or competing platforms is decreasing in the cost of information.*

Proof. The proof follows directly from the supermodularity property of all objective functions in $(c, -\delta)$. ■

This result is intuitive: when the cost of information increases, players react by reducing the amount of information disclosed to the advertisers. Figure 6 shows an example for $n = 20$. We observe that for low costs of information, the amount of information disclosed by the social planner, a monopoly platform and competing platforms coincide (and is maximum: $c = C = 10$). For larger costs of information disclosure, the amount of information is decreasing and the amount chosen by the social planner, the monopoly platforms and competing platforms diverge.

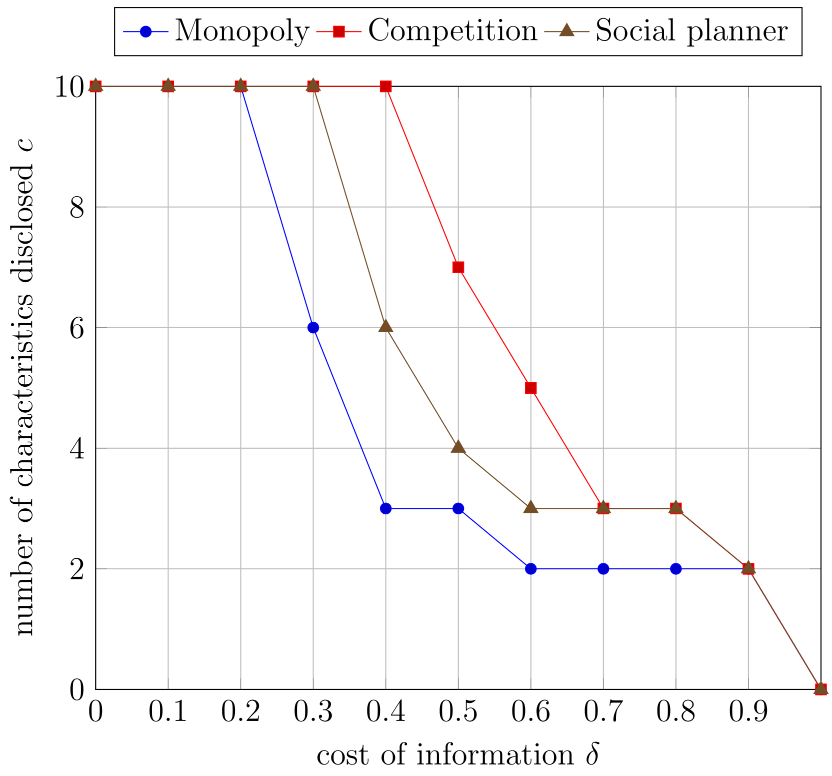

Figure 6: Impact of the cost of information on the amount of information disclosed, with $n = 20$.

5 Conclusion

This paper has discussed the value of consumer data in online advertising in a simple auction framework based on Ganuza (2004) and Ganuza and Penalva (2010), in which consumer data are formalized as a vector of characteristics that can be revealed to advertisers. This model is particularly relevant to better understand the role of consumer data in real time bidding systems, which are widely used to sell display ads online. This topic is at the heart of a recent sector-specific investigation into online advertising by the French competition authority (FCA (2018)).

The result that platforms with strong market power are less prone to use consumer data than what is socially optimal suggests that intervention to limit data collection by platforms may be misplaced. On the other hand, competition among selling platforms generate stronger incentives to collect data and may even lead to too strong a use of personal data. So, in some sense, data collection by one strong dominant platform may be less of a concern than when consumers use different platforms, which compete to attract advertisers. At the same time, profit maximizing platforms, irrespective of their market power, may face at least locally increasing returns to consumer data, which may lead to more concentration in the platform market.

Although it enables a rich treatment of information and consumer data, our model remains very simple on many aspects. One important limitation is our model is the assumption that advertisers single-home and coordinate their decision to join the same platform. A relevant question would be whether such a coordinated outcome could arise as an equilibrium of a richer model. In addition, data can be collected from many different sources, on many different platforms, and these data can then be processed, consolidated and aggregated by data brokers. So, another relevant development of our research would allow for such a formalization of the market for data and would lead to a more realistic representation of the current evolution of the digital economy. This extensions, however, are left for future research.

References

Acquisti, A., Taylor, C. and L. Wagman (2016). “The Economics of Privacy.” *Journal of Economic Literature*, 54(2), 442-492.

Armstrong, M. (2006). “Competition in two-sided markets.” *RAND Journal of Economics*, 37(3), 668-691.

Autorité de la Concurrence and Bundeskartellamt (2016). “Competition law and data.”

Board, S. (2010). “Revealing Information in Auctions: The Allocation Effect.” *Economic Theory* , 38(1), 125-135.

De Corniere, A. and R. de Nijs (2016). “Online advertising and privacy.” *RAND Journal of Economics*, 47(1), 48-72.

French Competition Authority (2018). “Opinion 18-A-03 of 6 March 2018 regarding data usage in the online advertising sector and its summary”

Federal Trade Commission (2014). “Data Brokers: A Call for Transparency and Accountability.”

Ganuza, J.J. (2004). “Ignorance Promotes Competition: An Auction Model with 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.

Lambrecht, A., Goldfarb, A., Bonatti, A., Ghose, A., Goldstein, D.G., Lewis, R., Rao, A., Sahni, N., and S. Yao (2014). “How do firms make money selling digital goods online?” *Marketing Letters*, 25(3), 331-341.

Office of Fair Trading (2010). “Online targeting of advertising and prices: A market study.” OFT 1231.

Office of Fair Trading (2013). “Personalised pricing - increasing transparency to improve trust.” OFT 1489.

Troncoso-Valverde, C. (2017). “Releasing information in private-value second-price auctions.” *Economic Theory* ,1-37.