Cowles Foundation

for Research in Economics at Yale University

Cowles Foundation Discussion Paper No. 1920R

SELLING COOKIES

Dirk Bergemann and Alessandro Bonatti

October 2013
Revised November 2013

Selling Cookies*

Dirk Bergemann[†] Alessandro Bonatti[‡]

November 1, 2013

Abstract

We analyze data pricing and targeted advertising. Advertisers seek to tailor their spending to the value of each consumer. A monopolistic data provider sells cookies–informative signals about individual consumers' preferences. We characterize the set of consumers targeted by the advertisers and the optimal monopoly price of cookies. The ability to influence the composition of the targeted set provides incentives to lower prices. Thus, the price of data decreases with the reach of the database and increases with the fragmentation of data sales. We characterize the optimal policy for selling information and its implementation through nonlinear pricing of cookies.

KEYWORDS: Data Providers, Information Sales, Targeting, Online Advertising, Media Markets.

JEL CLASSIFICATION: D44, D82, D83.

*We would like to thank Bob Gibbons, Michael Grubb, Richard Holden, Duncan Simester, Andy Skrzypacz, David Soberman, K. Sudhir, Juuso Toikka, Mike Whinston, and Jidong Zhou, as well as participants in various seminars and conferences for helpful discussions.
[†]Yale University, 30 Hillhouse Ave., New Haven, CT 06520, USA, `dirk.bergemann@yale.edu`
[‡]MIT Sloan School of Management, 100 Main Street, Cambridge MA 02142, USA, `bonatti@mit.edu`

1 Introduction

1.1 Motivation

The use of individual-level information is rapidly increasing in many economic and political environments, from advertising (various forms of targeting) to health insurance (contact and treatment of patients based on health histories) to electoral campaigns (identifying voters who are likely to switch or to turn out). In all these environments, the socially efficient match between individual and "treatment" may require the collection, analysis and diffusion of highly personalized data.

A large number of important policy and regulatory questions are beginning to emerge around the use of personal information. However, to properly frame these questions, we must first understand how markets for personalized information impact the creation and distribution of surplus, which is the main objective of this paper.

Much of the relevant data is collected and distributed by data brokers and data intermediaries ranging from established companies such as Acxiom and Bloomberg, to more recently established companies such as Bluekai and eXelate. Perhaps the most prevalent technology to enable the collection and resale of individual-level information is based on cookies and related means of recording browsing data. Cookies are small files placed by a website in a user's web browser that record information about the user's visit. Data providers use several partner websites to place cookies on user's computers and collect information. In particular, the first time any user visits a partner site (*e.g.*, a travel site), a cookie is sent to her browser, recording any action taken on the site during that browsing session (*e.g.*, searches for flights).[1] If the same user visits another partner website (*e.g.*, an online retailer), the information contained in her cookie is updated to reflect the most recent browsing history.

The data provider therefore maintains a detailed and up-to-date profile for each user, and compiles segments of consumer characteristics, based on each individual's browsing behavior. The demand for such highly detailed, consumer-level information is almost entirely driven by advertisers, who wish to tailor their spending and their campaigns to the characteristics of each consumer, patient, or voter.

The two distinguishing features of online markets for data are the following: (a) individual queries (as opposed to access to an entire database) are the actual products for sale,[2] and (b)

[1] This type of cookie is known as *third-party cookie*, because the domain installing it is different from the website actually visited by the user. Over half of the sites examined in a study by the Wall Street Journal installed 23 or more third-party cookies on visiting users' computers. (*The Web's New Gold Mine: Your Secrets,* the Wall Street Journal, July 30, 2010.) For a detailed report on "The State of Data Collection on the Web," see the 2013 Krux Cross Industry Study at http://www.krux.com/pro/broadcasts/krux_research/CIS2013/.

[2] For example, Bluekai and eXelate sell thousands of demographic, behavioral and purchase-intent seg-

linear pricing is predominantly used.[3] In other words, advertisers specify which consumer segments and how many total users ("uniques") they wish to acquire, and pay a price proportional to the number of users. These features are prominent in the market for cookies, but are equally representative of many markets for personal information.

In all these markets, a general picture emerges where an advertiser acquires very detailed information about a segment of "targeted" consumers, and is rather uninformed about a larger "residual" set. This kind of information structure, together with the new advertising opportunities, poses a number of economic questions. How is the advertisers' willingness to pay for information determined? Which consumers should they target? How should a data provider price its third-party data? How does the structure of the market for data (*e.g.*, competition among sellers, data exclusivity) affect the equilibrium price of information? More specifically to online advertising markets, what are the implications of data sales for the revenues of large publishers of advertising space?

In this paper, we explore the role of data providers on the price and allocation of consumer-level information. We provide a framework that addresses general questions about the market for data, and contributes to our understanding of recent practices in online advertising. Thus, we develop a simple model of data pricing that captures the key trade-offs involved in selling the information encoded in third-party cookies. However, our model also applies more broadly to markets for consumer-level information, and it is suited to analyze several offline channels as well.

In our model, we consider heterogeneous consumers and firms. The (potential) surplus is given by a function that assigns a value to each realized match between a consumer and a firm (the match value function). The match values differ along a purely horizontal dimension, and may represent a market with differentiated products and heterogeneous preferences of consumers over the product space. In order to realize the potential match value, each firm must "invest" in contacting consumers. An immediate interpretation of the investment decision is advertising spending that generates contacts and eventually sales. We refer to the advertising technology as the rate at which investment into contacts generates actual sales.

We maintain the two distinguishing features of selling cookies (individual queries and per-user "bit" pricing) as the main assumptions. These assumptions can be stated more precisely

ments. See http://exelate.com/wp-content/uploads/2012/02/Data-Segments-Overview-013112b.pdf for an example of the segments for sale.

[3] Cookie-based data can be priced in two ways: *per stamp* (CPS), where buyers pay for the right to access the cookie, independent of number of uses of the data; and *per mille* (CPM), where the price of the data is proportional to the number of advertising impressions shown using that data. Most data providers give buyers a choice of the pricing criterion. (*Data with Benefits*, Online Metrics Insider, October 25, 2010. http://www.mediapost.com/publications/article/158198/#axzz2Z7WyhSoj.)

as follows: (a) *individual queries are for sale:* We allow advertisers to purchase information on individual consumers. This enables advertisers to segment users into a targeted group that receives personalized levels of advertising, and a residual group that receives a uniform level of advertising (possibly zero). More formally, this means the information structures available to an advertiser are given by specific partitions of the space of match values. (b) *data bits are priced separately*: We restrict the data provider to set a uniform unit price, so that the payment to the data provider is proportional to the number of users ("cookies") acquired.

There exist, of course, other ways to sell information, though linear pricing of cookies is a natural starting point. We address these variations in extensions of our baseline model. In particular, we explore alternative mechanisms for selling information, such as bundling and nonlinear pricing of data. We formally establish conditions under which linear pricing provides a good approximation for the optimal mechanism.

1.2 Overview of the Results

In Section 3, we characterize the advertisers' demand for information for a given price of data. We establish that advertisers purchase information on two convex sets of consumers, specifically those with the highest and lowest match values. Intuitively, advertisers will not buy information about every consumer. Advertisers must then estimate the match value within the residual set of consumers, and excluding a convex set allows them to minimize the prediction error. Furthermore, under quite general conditions, the data-buying policy takes the form of a single cutoff match value. More surprisingly, advertisers may buy information about all users above the cutoff value (*positive targeting*) or below the cutoff value (*negative targeting*). Each of these data-buying policies alleviates one potential source of advertising mismatch, namely wasteful spending on low-value matches and insufficient intensity on high-value matches. The optimality of positive vs. negative targeting depends on the advertising technology and on the distribution of match values, *i.e.*, on properties of the complete-information profit function alone.

In Section 4, we turn to the data provider's pricing problem. We first examine the subtle relationship between the monopoly price and the unit cost of advertising. The cost of advertising reduces both the payoff advertisers can obtain through better information, and their payoff if uninformed. The overall effect on the demand for cookies and on the monopoly price is, in general, non-monotone. In an informative example, the monopoly price of cookies is single-peaked in the cost of advertising. This suggests which market conditions are more conducive to the profitability of a data provider.

We then examine the role of market structure on the price of cookies. Within our monopoly framework, we explore the possibility of consumers selling their own information. Formally, we consider a continuum of information providers, each one selling one signal exclusively. Surprisingly, we find that concentrating sales in the hands of a single data provider is not necessarily detrimental to social welfare, and that prices are higher under fragmentation. The reason for this result is that exclusive sellers are not really competing in prices. On the contrary, they ignore the negative externality that raising the price of one signal imposes on the advertisers' demand for information about all other consumers. A similar mechanism characterizes the effects of an incomplete database, sold by a single firm. In that case, the willingness to pay for information increases with the size of the database, but the monopoly price may, in fact, decrease.

In Section 5, we enrich the set of pricing mechanisms available to the data provider. In particular, in a binary-action model, we introduce nonlinear pricing of information structures. We show that the data provider can screen vertically heterogeneous advertisers by offering subsets of the database at a decreasing marginal price. The optimal nonlinear price determines exclusivity restrictions on a set of "marginal" cookies: in particular, second-best distortions imply that some cookies that would be profitable for several advertisers are bought by a subset of high-value advertisers only.

Finally, in Section 6, we examine the interaction between the markets for data and online advertising. In particular, we relate the properties of the advertising technology to the payoff externalities that the price in one market imposes on the seller in the other market. In addition, these properties determine the publisher's incentives to acquire information and to release it to the advertisers. A consistent pattern emerges linking the advertisers' preferences for positive vs. negative targeting and the degree to which a publisher wishes to improve the targeting opportunities available to them.

1.3 Related Literature

The issue of optimally pricing information in a monopoly and in a competitive market has been addressed in the finance literature, starting with seminal contributions by Admati and Pfleiderer (1986), Admati and Pfleiderer (1990) and Allen (1990), and more recently by García and Sangiorgi (2011). A different strand of the literature has examined the sale of information to competing parties. In particular, Sarvary and Parker (1997) model information-sharing among competing consulting companies; Xiang and Sarvary (2013) study the interaction among providers of information to competing clients; Iyer and Soberman (2000) analyze the sale of heterogeneous signals, corresponding to valuable product modifications, to firms competing in a differentiated-products duopoly; Taylor (2004) studies the sale of consumer lists that facilitate price discrimination based on purchase history. All of these earlier papers only allowed for the complete sale of information. In other words, they focused on signals that revealed (noisy) information about all realizations of a payoff-relevant random variable. The main difference with our paper's approach is that we focus on "bit-pricing" of information, by allowing a seller to price each realization of a random variable separately.

The literature on the optimal choice of information structures is rather recent. Bergemann and Pesendorfer (2007) consider the design of optimal information structures within the context of an optimal auction. There, the principal controls the design of both the information and the allocation rule. More recently, Kamenica and Gentzkow (2011) consider the design of the information structure by the principal when the agent will take an independent action on the basis of the received information. Rayo and Segal (2010) examine a similar question in a model with multidimensional uncertainty and private information on the agent's cost of action. In our model, the advertisers' demand for data is reminiscent of information acquisition under the rational inattention problem, as in Sims (2003). The main difference with both the inattention and persuasion literature is that we endogenize the agent's information cost parameter by explicitly analyzing monopoly pricing, rather than directly choosing an information structure.

In related contributions, Anton and Yao (2002), Hörner and Skrzypacz (2012), and Babaioff, Kleinberg, and Paes Leme (2012) derive the optimal mechanism for selling information about a payoff-relevant state, in a principal-agent framework. Anton and Yao (2002) emphasize the role of partial disclosure; Hörner and Skrzypacz (2012) focus on the incentives to acquire information; and Babaioff, Kleinberg, and Paes Leme (2012) allow both the seller and the buyer to observe private signals. Finally, Hoffmann, Inderst, and Ottaviani (2013) consider targeted advertising as selective disclosure of product information to consumers with limited attention spans.

The role of specific information structures in auctions, and their implication for online advertising market design, are analyzed in recent work by Abraham, Athey, Babaioff, and Grubb (2012), Celis, Lewis, Mobius, and Nazerzadeh (2012), and Kempe, Syrganis, and Tardos (2012). All three papers are motivated by asymmetries in bidders' ability to access additional information about the object for sale. Finally, Mahdian, Ghosh, McAfee, and Vassilvitskii (2012) study the revenue implications of cookie-matching from the point of view of an informed seller of advertising space, uncovering a trade-off between targeting and information leakage.

In our earlier work, Bergemann and Bonatti (2011), we analyzed the impact that exogenous changes in the information structures have on the competition for advertising space.

In the present strategic environment, the pricing decision of the data provider and the data purchasing decision of the advertiser endogenously determined the information structure and hence the equilibrium level of targeting and advertising.

2 Model

2.1 Matching and Preferences

We consider a unit mass of consumers (or "users"), $i \in [0,1]$, and firms (or "advertisers"), $j \in [0,1]$, a single publisher, and a monopolistic data provider. The consumers and firms are each uniformly distributed on the unit interval. Each consumer-firm pair (i,j) generates a (potential) match value $v : [0,1] \times [0,1] \to V$, with $V = [\underline{v}, \bar{v}] \subseteq \mathbb{R}_+$.

The (uniform) distribution over the consumer-firm pairs (i,j) generates a distribution of values through the match value function v. For every measurable subset A of values in V, we denote the resulting measure by μ:

$$\mu(A) \triangleq \int_{\{i,j \in [0,1] | v(i,j) \in A\}} \mathrm{d}i\mathrm{d}j.$$

Consider the set of matches that generates a value v or less,

$$A^v \triangleq \{i,j \in [0,1] \, | v(i,j) \leq v\}.$$

The associated distribution function $F : V \to [0,1]$ is defined by

$$F(v) \triangleq \mu(A^v).$$

By extension, we define the *conditional* measure for every consumer i and every firm j by:

$$\mu_i(A) \triangleq \int_{\{j \in [0,1] | v(i,j) \in A\}} \mathrm{d}j, \quad \text{and} \quad \mu_j(A) \triangleq \int_{\{i \in [0,1] | v(i,j) \in A\}} \mathrm{d}i,$$

and associated conditional distribution functions $F_i(v)$ and $F_j(v)$. We assume that the resulting match values are identically distributed across consumer and across firms, *i.e.*, for all i, j, and v:

$$F_i(v) = F_j(v) = F(v).$$

Prominent examples of distributions that satisfy our symmetry assumption include: i.i.d. match values across consumer-firm pairs; and uniformly distributed firms and consumers around a unit-length circle, where match values are a function of the distance $|i-j|$.

Thus, match values differ along a purely horizontal dimension. This assumption captures the idea that, even within an industry, the same consumer profile can represent a high "match value" to some firms and at the same time a low "match value" to others. This is clearly true for consumers that differ in their geographical location, but applies more broadly as well. Consider the case of credit-score data: major credit card companies are interested in reaching consumers with high credit-worthiness; banks that advertise consumer credit lines would like to target individuals with average scores, who are cash-constraint, but unlikely to default; and subprime lenders such as used car dealers typically cater to individuals with low or non-existing credit scores.[4]

Firm j must take an action $q_{ij} \geq 0$ directed at consumer i to realize the potential match value $v(i,j)$. We define q as the *match intensity*. We abstract from the details of the revenue-generating process associated to matching with intensity q. The complete-information profits of a firm generating a contact of intensity q with a consumer of value v are given by

$$\pi(v,q) \triangleq vq - cm(q). \tag{1}$$

The *matching function* $m : \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be increasing, continuously differentiable, and convex. In the context of advertising, q corresponds to the probability of generating consumer i's awareness about firm j's product. If consumer i is made aware of the product, he generates a net present value to the firm equal to $v(i,j)$. Awareness is generated by buying advertising space from the publisher, and we assume that advertising space can be purchased at a constant marginal cost c.

The *advertising technology* is then summarized by the matching function $m(q)$ that represents the amount of advertising space m required to generate a match with probability q and the marginal cost c of advertising space. With the advertising application in mind, we may view q as scaling the consumer's willingness to pay directly, or as the amount of advertising effort exerted by the firm, which also enters the consumer's utility function. Thus, the profit function in (1) is consistent with the informative, as well as the persuasive and complementary views of advertising (see Bagwell, 2007).

2.2 Information and Timing

Initially, neither the advertisers nor the publisher have information about the pair-specific match values $v(i,j)$ beyond the common prior distribution $F(v)$. Each advertiser can purchase information from a monopolist data provider to better target his advertising choices.

[4]See Adams, Einav, and Levin (2009) for a description and model of subprime lending.

The data provider has information relating each consumer to a set of characteristics, represented by the index i. Advertisers can then query the data provider for the "user IDs" of consumers with specific characteristics i. From the perspective of advertiser j, the only relevant aspect of the characteristic of consumer i is the value of the interaction $v(i,j)$. Thus, if advertiser j wishes to identify (and contact) all consumers with valuation v, then he requests the identity of all consumers with characteristics i such that $v = v(i,j)$.

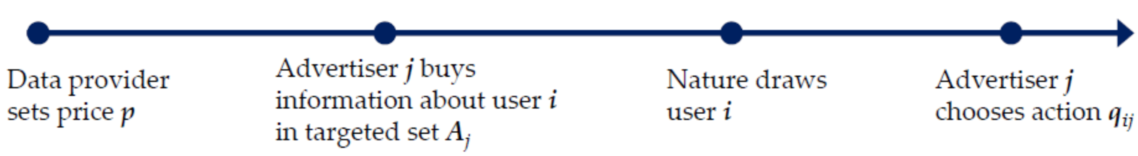
Each advertiser j can purchase information about any subset of consumers with given set of match values $A_j \subset V$. We shall hereafter refer to "cookie v" as the characteristics of i that identify, for advertiser j, all consumers such that $v(i,j) = v$. Thus, if firm j purchases cookie v, then the value $v = v(i,j)$ belongs to the set A_j. Advertisers are able to tailor their action q to each consumer they acquire cookies about. For this reason, we refer to the sets A_j and A_j^C as the *targeted set* and the *residual set* (or complementary set), respectively.

We assume that the data about individual consumer is sold at a constant linear price per cookie. Thus the price of the targeted set A_j is given by

$$p(A_j) \triangleq p\mu(A_j). \tag{2}$$

This assumption reflects the pricing of data "per unique user" (also known as "cost per stamp"). It also matches the offline markets for data, where the price of mailing lists, or lists of credit scores is related to the number of user records, and where the data cannot be bought contextually to its use. Figure 1 summarizes the timing of our model.

FIGURE 1: TIMING



3 Demand for Information

When facing consumer i, each advertiser chooses the advertising intensity q_{ij} to maximize profits, given the expectation of $v(i,j)$. The value of information for each advertiser is determined by the incremental profits they could accrue by purchasing more cookies. For the consumers in the targeted set A_j, advertiser j is able to perfectly tailor his advertising spending to all consumers included in the targeted set A_j. In particular, we denote the complete-information demand for advertising space $q^*(v)$ and profit level $\pi(v)$ by

$$\begin{aligned} q^*(v) &\triangleq \arg\max_{q\in\mathbb{R}_+}[\pi(v,q)], \\ \pi(v) &\triangleq \pi(v,q^*(v)). \end{aligned}$$

In contrast, each advertiser j must choose a constant level of q for all consumers in the complement (or residual) set A_j^{C}. The optimal level of advertising $q^*(A_j^{\mathrm{C}})$ depends in principle on the targeted set A_j. Because the objective $\pi(v,q)$ is linear in v, it is given by

$$q^*(A_j^{\mathrm{C}}) \triangleq \arg\max_{q\in\mathbb{R}_+}[\mathbb{E}[\pi(v,q) \mid v \notin A_j]] = q^*(\mathbb{E}[v \mid v \notin A_j]).$$

Therefore, we can formulate each advertiser's information-acquisition problem as the choice of a measurable subset A of the set of match values V:

$$\max_{A\subseteq V}\left[\int_A(\pi(v,q^*(v))-p)\,\mathrm{d}F(v)+\int_{A^{\mathrm{C}}}\pi(v,q^*(A^{\mathrm{C}}))\mathrm{d}F(v)\right], \tag{3}$$

where by symmetry we can drop the index j for the advertiser.

By including all consumers with match value v into the targeted set A, the advertiser can improve his gross profits from the uninformed level to the informed level, albeit at the unit cost p per consumer. In problem (3), the total price paid by the advertisers to the data provider is then proportional to the measure of the targeted set.

The program formulated in (3) is reminiscent of a rational inattention problem as in Sims (2003). To be clear, we are not imposing a constraint on any player's total attention span. Instead, the limits to attention here are due to a direct monetary cost p, which reduces the advertisers' incentives to acquire perfect information about all match values. Problem (3) also shares some features with the Bayesian persuasion literature, *e.g.*, Rayo and Segal (2010) and Kamenica and Gentzkow (2011). As in those models, one party (here, the advertiser) chooses an information structure to maximize a given objective. Another party (here, the data provider) would like the chosen information structure to be as precise as possible: in Kamenica and Gentzkow (2011), in order to take a subsequent action; and in our paper, to maximize profits. In both cases, a wedge (or bias) makes the two parties' ideal information structures differ. In this paper, the wedge is represented by the price of data. For a fixed p, the two models differ because the data provider (our receiver) does not take any action. However, when we analyze the pricing problem in Section 4, we allow the data provider to take an action *before* the advertiser (our sender) chooses an information structure. Thus, with reference to both the inattention and persuasion literatures, the greatest difference with this paper is that we allow the data provider to choose the information cost (preference bias).

We now seek to characterize the properties of the optimal targeted set, as a function of the unit price of cookies p and of the cost of acquiring advertising space c. We begin with a simple example.

3.1 The Binary-Action Environment

We start with linear matching costs and uniformly distributed match values; we then generalize the model to continuous actions and general distributions. Formally, let $F(v) = v$, with $v \in [0,1]$ and $c \cdot m(q) = c \cdot q$, with $q \in [0,1]$. The linear cost assumption is equivalent to considering a binary action environment, $q \in \{0,1\}$, as the optimal contact policy will always be to choose $q \in \{0,1\}$.

In this simplified version of the model, targeting is very coarse: under complete information, it is optimal to contact a consumer v (*i.e.*, to choose $q^*(v) = 1$) if and only if the match value v exceeds the unit cost of advertising c. Thus, the complete-information profits are given by

$$\pi(v) \triangleq \max\{v - c, 0\}. \tag{4}$$

Likewise, the optimal action on the residual set is given by

$$q^*(A^{\mathrm{C}}) = 1 \iff \mathbb{E}[v \mid v \in A^{\mathrm{C}}] \geq c.$$

In this binary setting, advertisers always choose a *constant* action $q \in \{0,1\}$ on the targeted set A and a *different constant* action on the residual set A^{C}: intuitively, information about consumer v has positive value only if it affects the advertiser's subsequent action. Therefore, advertisers adopt one of two mutually exclusive strategies to segment the consumer population: (*i*) *positive targeting* consists of buying information on the highest-value consumers, contacting them and excluding everyone else; (*ii*) *negative targeting* consists of buying information on the lowest-value consumers, avoiding them and contacting everyone else.

Consider the willingness to pay for the marginal cookie under each targeting strategy. If the advertiser adopts positive targeting, then he purchases information on (and contacts) all consumers above a certain threshold. Conversely, if the advertiser adopts negative targeting, then he purchases all the cookies below a certain threshold. The optimality of a threshold strategy follows from the monotonicity of the profit in v and the binary action environment. The choice of the optimal targeting strategy and the size of the targeted set naturally depends on the cost of contact c and on the price of information p. We denote the optimal targeted set by $A(c,p)$. This set is defined by a threshold value v^* that either determines a lower interval $[\underline{v}, v^*]$ or an upper interval $[v^*, \bar{v}]$, depending on the optimality of either negative or positive targeting, respectively.

In the binary environment, we can explicitly identify the size of the targeted set. If the advertiser adopts positive targeting, then he purchases information on all consumers up to the threshold v^* that leaves him with nonnegative net utility, or $v^* = c + p$. Conversely, if the advertiser adopts negative targeting, then at marginal cookie, the gain from *avoiding* the contact, and thus saving $c - v$, is just offset by the price p of the cookie, and thus $v^* = c - p$. Under either targeting strategy, the advertiser trades off the magnitude of the error made on the residual set with the cost of acquiring additional information.

The cost of the advertising space, the matching cost c, determines whether positive or negative targeting is optimal. If c is high, only a small number of high value users are actually profitable. For any price p of information, it is then optimal for advertisers to buy only a small number of cookies and to contact only those with very high values. The opposite intuition applies when the cost of contact c is very low: almost all users are profitable, advertisers only buy a few low valuation cookies v, and exclude the corresponding users.

Proposition 1 (Targeting Strategy)
For all $c, p > 0$, the optimal targeted set $A(c,p)$ is given by:

$$A(c,p) = \begin{cases} [0, \max\{c-p, 0\}] & \text{if} \quad c < 1/2; \\ [\min\{c+p, 1\}, 1] & \text{if} \quad c \geq 1/2. \end{cases}$$

Proposition 1 establishes that the residual and the targeted set are both connected sets (intervals), and that advertisers do not buy information about all consumers. The first result is specific to the binary action environment. The latter is a more general implication of inference about the values in the residual set from the shape of the targeted set. After all, if the advertiser were to buy (almost) all cookies, then he might as well reduce his cookie purchases on a small interval, save the corresponding cookie costs, while still selecting the correct action on the residual set.[5]

The binary environment illustrates some general features of optimal targeting and information policies. In particular, three implications of Proposition 1 extend to general settings: (a) the residual set is non-empty; (b) advertisers do not necessarily buy the cookies of high value consumers; and (c) the cost c of the advertising space guides their strategy. At the

[5]The value $c = 1/2$ of the threshold which determines the choice of targeting strategy happens to coincide with the threshold value that would determine whether advertisers contact all consumers, or none, in the absence of cookies. This is a special feature of the uniform distribution.

same time, the binary environment cannot easily capture several aspects of the model, including the following: the role of the distribution of match values (and of the relative size of the left and the right tail in particular); the role of precise tailoring and the need for more detailed information; the determinants of the advertisers' optimal targeting strategy; and the effect of the cost of advertising on the demand for information.

3.2 The Continuous-Action Environment

We now proceed to analyze the general version of our model, in which we consider a continuum of actions and a general distribution of match values. It is helpful to first describe the demand for advertising space when the value of the match v is known to the advertisers. Thus, we introduce the *complete-information* decision and profits. As specified earlier, we allow for a general differentiable, increasing and convex cost function $m(q)$. We shall further assume that $m'(0) = 0$, which implies that the complete-information demand for advertising is positive for all match values.

The complete-information demand for advertising space, denoted by $q^*(v)$, is characterized by the first-order condition

$$v = cm'(q^*(v)). \tag{5}$$

By contrast, if the advertiser has access to the distribution $F(v)$ only, the *prior-information* demand for advertising space $\bar{q}$ is given by

$$\bar{q} \triangleq q^*(A^{\mathrm{C}} = V) = q^*(\mathbb{E}[v]).$$

More generally, given a targeted set A, the optimal advertising level on the residual set A^{C} satisfies the following condition:

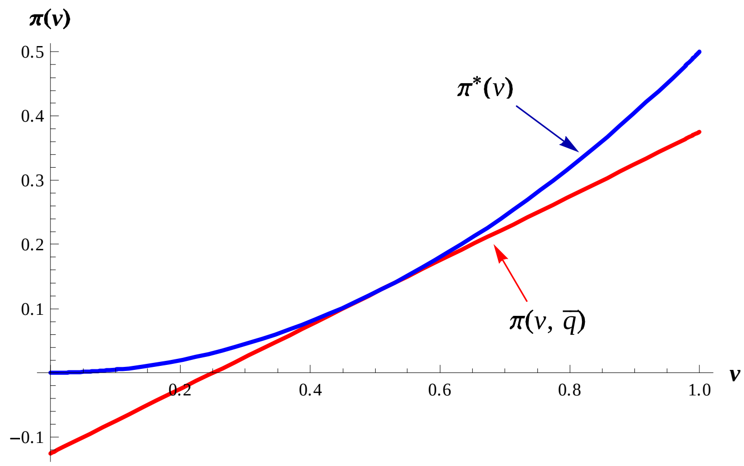
$$\mathbb{E}[v \mid v \in A^{\mathrm{C}}] = cm'(q^*(A^{\mathrm{C}})). \tag{6}$$

Thus, our continuous-action model has the two key features that advertisers (a) differentiate spending levels within the targeted set, and (b) choose a uniform (strictly positive) advertising level for the residual set. In turn, the optimal advertising level on the residual set $q^*(A^{\mathrm{C}})$ varies with the composition of the targeted set A. These advertising policies might arguably represent the choices of a large brand marketer who wishes to fine-tune spending on a group of consumers, while adopting "umbrella spending" on everyone else.

Note that the realized *complete-information* profit $\pi^*(v)$ is strictly convex in v, as $q^*(v)$ is strictly increasing, and the objective function $\pi(v, q)$ is linear in v. In contrast, the realized profit under *prior information* from a consumer with value v is linear in v, and it is given by

$\pi(v,\bar{q})$. Figure 2 describes the profit function under complete information $\pi^*(v)$ and prior information $\pi(v,\bar{q})$.

Figure 2: Complete- and Prior-Information Profits



As intuitive, under prior information, the firm chooses excessive (wasteful) advertising to low-value consumers and insufficient advertising to higher-value consumers. The firm therefore has a positive willingness to pay for information, *i.e.*, for cookies. The value of information for every match value v is visually described by the difference between the complete information and the prior information profit function:

$$\pi^*(v) - \pi(v,\bar{q}). \tag{7}$$

Figure 2 suggests that the value of information is highest for extreme match values.[6] Consequently, the next result establishes the optimality of a convex *residual set* of cookies. Each advertising firm purchases all cookies in a set $A = [\underline{v}, v_1] \cup [v_2, \bar{v}]$. The value of the lower and upper threshold are determined again by c and p, thus $v_1 \triangleq v_1(c,p)$ and $v_2 \triangleq v_2(c,p)$, respectively. Proposition 2 confirms the intuition that the value of information is lowest for intermediate match values and highest for match values on the tails.

Proposition 2 (Convexity of Residual Set)
For all $c, p > 0$, the optimal residual set $A^C(c,p)$ is a non-empty interval $[v_1(c,p), v_2(c,p)]$.

Proposition 2 allows us to rewrite the firm's problem (3) as the choice of two thresholds,

[6] In this example, $cm(q) = q^2/2$, and $F(v) = v$, $v \in [0,1]$.

v_1 and v_2, that define the targeted and residual sets:

$$\max_{v_1,v_2} \int_{v_1}^{v_2} \left[p + \pi\left(v, q^*\left([v_1, v_2]\right)\right) - \pi\left(v\right)\right] \mathrm{d}F\left(v\right), \tag{8}$$
$$\text{s.t. } cm'\left(q^*\left([v_1, v_2]\right)\right) = \mathbb{E}\left[v \mid v \in [v_1, v_2]\right].$$

In program (8), as the bounds of the residual set are stretched (*e.g.*, as v_1 decreases), the advertiser earns a marginal benefit of p and incurs a marginal cost of $\pi\left(v_1\right) - \pi\left(v_1, q^*([v_1, v_2])\right)$. In addition, the advertiser adjusts the optimal action on the residual set to take the new inference problem into account. (Of course, this has no first-order effect on profits at the optimum.) The average match value $\mathbb{E}\left[v \mid v \in [v_1, v_2]\right]$ determines the demand for advertising space in the residual set $q^*\left([v_1, v_2]\right)$, which in turn affects the value of information.

In the discussion above, we described the value of information above as the difference between the profit of an informed and an uninformed advertiser $\pi^*\left(v\right) - \pi\left(v, \bar{q}\right)$. This revenue comparison is conditional on the realization of the value v, and it is thus an ex-post comparison. For the complete determination of the optimal policy, the advertiser has to evaluate how large these gains from information are from an *ex-ante* point of view. The advertiser therefore has to weigh the likelihood of different realizations, represented by the distribution $F\left(v\right)$ of values, and the gains from responding to the information, represented by the convexity of the matching function $m\left(q\right)$. To understand the exact nature of these trade-offs, it is useful to begin with a "symmetric" environment for $F\left(v\right)$ and $m\left(q\right)$. In the context of negative vs. positive targeting, this corresponds to a symmetric distribution $F\left(v\right)$ around the mean $\mathbb{E}\left[v\right]$ and a quadratic matching function $m\left(q\right)$, such as in the example of Figure 2.

3.2.1 Joint Targeting: Positive and Negative

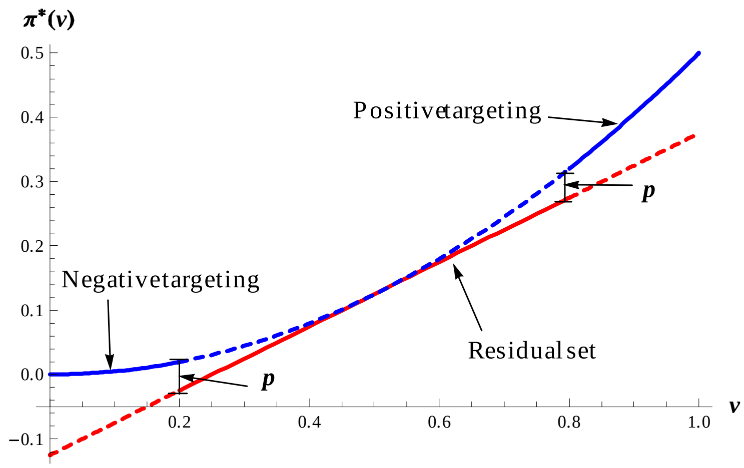
When matching costs are quadratic and match values are symmetrically distributed, advertisers always choose to target both low- and high-valuation consumers. In addition, under these symmetry conditions, the residual set (*i.e.*, the set of excluded valuations) is an interval centered on the prior mean $\mathbb{E}\left[v\right]$. With a quadratic matching function, the optimal complete information matching intensity is linear in v, or $q^*\left(v\right) = v/c$. Moreover, the *gains from information* relative to the optimal matching policy for the mean value $q^*\left(\mathbb{E}\left[v\right]\right)$ are identical for values equidistant from the mean, regardless of whether they are below or above the mean. Of course, the value of information arises from adjustments of the matching intensity relative to the mean, *i.e.*, increasing the matching intensity for values above the mean and decreasing the matching intensity for values below the mean. Furthermore, because the curvature of the cost function is constant in q when $m\left(q\right)$ is quadratic, this symmetry argument holds under any symmetric distribution $F(v)$. Proposition 3 verifies the above intuition.

Proposition 3 (Positive and Negative Targeting)
With symmetrically distributed match values and quadratic matching costs, the optimal residual set is given by:

$$A^C(c,p) = [\mathbb{E}[v] - 2\sqrt{cp},\ \mathbb{E}[v] + 2\sqrt{cp}].$$

The measure of the residual set is increasing in the product of the price of information p and the cost of contact c. Thus, an increase in either one depresses the number of cookies acquired, and shrinks the targeted set by expanding the residual set toward the tails of the distribution. Figure 3 illustrates the demand for cookies and the resulting profit levels in the quadratic environment.

Figure 3: Positive and Negative Targeting



The symmetry conditions introduced in Proposition 3 have important implications not only for the optimal location of the residual set, but also for its size. In particular, the expected match value in the residual set is equal to the prior mean $\mathbb{E}[v]$, regardless of the measure of the residual set A^C. Therefore, the quantity of signals purchased by the advertiser does not influence the uninformed action $\bar{q}$, and hence it does not affect the marginal value of information. This also implies that the willingness to pay for information about any consumer v is independent of the distribution of match values.

In turn, the interaction between the symmetric gains from information and the symmetry in the distribution suggest conditions under which either only positive or only negative targeting become optimal, as we establish in the next set of results.

3.2.2 Exclusive Targeting: Positive or Negative

While the residual set is always connected, as established by Proposition 2, the targeted set may be as well. In particular, the choice of a single (positive or negative) targeting policy depends on the value of information, and on its monotonicity properties over any interval. Proposition 4 establishes sufficient conditions under which firms demand cookies in a single interval, *i.e.*, they choose positive or negative targeting only.

Proposition 4 (Exclusive Targeting)

1. *If $m''(q)$ and $f(v)$ are decreasing, positive targeting is optimal:*

$$A(c,p) = [v_2(c,p), \bar{v}], \text{ and } v_2 > \underline{v}.$$

2. *If $m''(q)$ and $f(v)$ are increasing, negative targeting is optimal:*

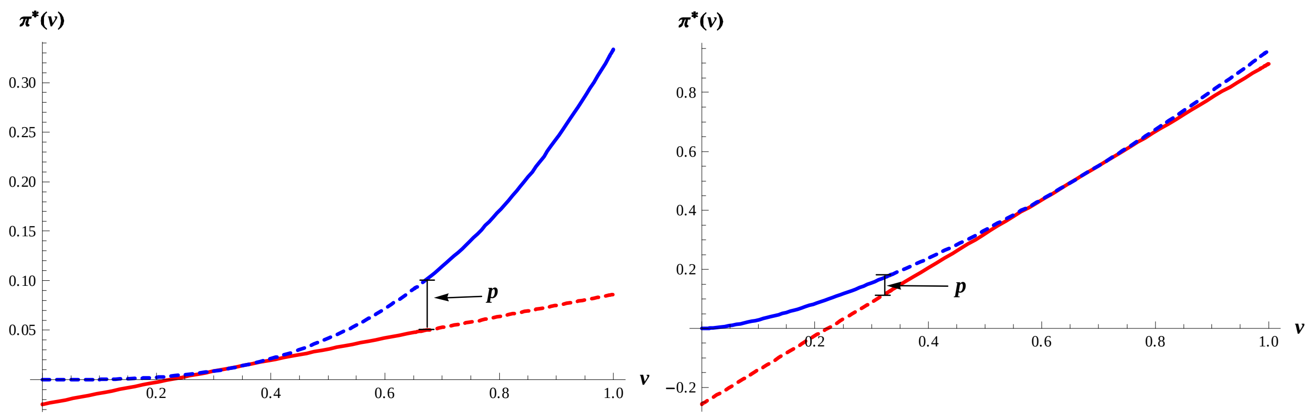
$$A(c,p) = [\underline{v}, v_1(c,p)], \text{ and } v_1 < \bar{v}.$$

The sufficient conditions in Proposition 4 for exclusive targeting are perhaps best understood when viewed as departures from the symmetric conditions of Proposition 3. If, say, positive targeting is to dominate negative targeting, then it has to be the case that the gains from information are larger on the upside than on the downside of values. Recall that the gains from information given the realization v are equal to $\pi^*(v) - \pi(v, \bar{q})$. Thus, if the curvature of the matching function $m''(q)$ is decreasing, the gains from information for realizations v equidistant from the mean $\mathbb{E}[v]$ are larger above the mean than below. Now, this pairwise comparison and reasoning could be undone by the relative likelihood of these two events. Thus, for the sufficient conditions, we need to guarantee that the distribution of values supports this pairwise argument, and hence the corresponding monotonicity requirement on the density $f(v)$. Figure 4 shows the equilibrium profit levels under positive targeting (A) and negative targeting (B).[7]

The optimality of targeting consumers in a single interval can be traced back to the two sources of the value of information, *i.e.*, wasteful advertising for low types and insufficient advertising for valuable consumers. Proposition 4 relates the potential for mismatch risk to the properties of the match cost function. In particular, when the curvature of the matching cost function is increasing, it becomes very expensive to tailor advertising to high-value consumers. In other words, the risk of insufficient advertising is not very high, given the cost

[7] In both panels, $F(v) = v$, $v \in [0,1]$ and $m(q) = q^b/b$. In panel (A), $b = 3/2$, and in panel (B), $b = 3$.

Figure 4: Positive or Negative Targeting



(A) (B)

of advertising space. The firm then purchases cookies related to lower-valued consumers.[8] Finally, note that the distribution of match values must affect the targeting decision, as the advertiser trades off the amount of learning (related to the range of the residual set $|v_2 - v_1|$) with the cost of acquiring the information (related to the probability measure of the targeted set). After all, the less likely events require a smaller expense in terms of the cost of cookies.

3.3 Empirical Relevance

We conclude the section on the demand for information by discussing the relevance of positive- and negative-targeting strategies for online advertising. In practice, an advertiser may adopt either or both strategies, and the choice a strategy in any specific context depends on the distribution of consumer values and on the cost of advertising. For instance, in the market for credit scores, a credit card company may want to acquire the profiles of consumers with the lowest scores, and make sure *not* to reach out to them; or it may select a small group of high credit-worthiness consumers, and reach out to them more aggressively. In the context of retail shopping, Pancras and Sudhir (2007) document the widespread use of both positive and negative targeting by catalogue merchants and manufacturers, and study the pricing of data by several intermediaries.

More recent studies provide indirect evidence in favor of adopting negative targeting to exploit the consumers' purchase cycle. For example, in the context of sponsored-search

[8]Examples of matching cost functions with concave marginal costs include power functions, $m(q) = q^a$ with $a < 2$. Examples of convex marginal costs include those derived from the Butters (1977) exponential matching technology, *i.e.*, $m(q) = -a \ln(1-q)$, with $a > 0$, and power functions $m(q) = q^a$, with $a > 2$.

advertising, Blake, Nosko, and Tadelis (2013) document that eBay obtains a positive return on investment only for consumers who have not visited the eBay site in the last two months. A similar pattern for the profitability of different customers also appears in the case of (offline) direct-marketing companies documented by Anderson and Simester (2013). In both contexts, a cost-efficient strategy for retailers consists of acquiring information about consumers with recent purchases and appropriately reducing the amount of advertising directed at them. These consumers are both low-value (at this point in their purchase cycle) and low in number, relative to the overall population, which makes negative targeting especially profitable.[9]

Finally, as real-time bidding makes online data markets more integrated with the advertising exchanges, we can identify two contrasting forces in terms of our model. On the one hand, the combined sale of data and advertising favors positive targeting by construction. On the other hand, when the cost of the data is tied to the price paid for advertising, contacting high-value becomes increasingly costly. If targeting through cookies results in a higher marginal cost of advertising, advertisers may specify lower bids for selected consumer segments (*i.e.*, adopt negative targeting) in order to reduce their total expenditure.[10]

4 The Price of Data

We now turn our attention to the optimal price of data, and we examine several aspects of monopoly pricing for cookies. We begin with the role of the cost c of advertising in the determination of the price p of cookies, and then we move to the implications of fragmented data sales and incomplete databases. We highlight the role of the residual set in determining the willingness to pay for information, and of the ability of the monopolist to influence its composition.

4.1 Data and Advertising: Complements or Substitutes?

We first examine the relationship between the cost of advertising c and the optimal monopoly price of cookies. For now, we focus our attention on the response of the optimal price $p^*(c)$ to changes in the cost c of advertising. We shall analyze the interaction between the data provider and the publisher in more detail in Section 6.

From the point of view of the advertiser, the data provider and the publisher are part of

[9] While advertisers may be able to identify their own repeat shoppers, they need to purchase third-party information about their competitors' customers who are at a similar stage in their purchase cycle.

[10] Interestingly, when advertising and information are sold contextually, negative targeting is explicitly allowed as a refinement option by most large providers of advertising space, including Google, Yahoo!, and Facebook.

value chain to match advertiser and consumer. It is therefore tempting to view the interaction of the data provider and publisher as a vertical chain (formed by strategic complements), and to associate with it the risk of double marginalization. This would suggest that an increase in the cost c of advertising would lead optimally to a partially offsetting decrease in the price of information $p^*(c)$. But at closer inspection, the relationship between the price of data and that of advertising is more subtle.

The purchase of data may allow the advertiser to concentrate the purchase of advertising space on a smaller but highly relevant segment. Thus, from the point of view of the advertiser, the data provides an option whose value might be increasing as the advertising space becomes more expensive. Thus, the purchase of data acts as a strategic complement for the high value realizations, but as a strategic substitute for the low valuations. After all, after learning of a low value consumer, the advertiser lowers his matching intensity, and might even set it equal to zero.

Therefore, by necessity, data purchases act simultaneously as strategic substitutes as well as complements to advertising purchases. This subtlety in the interaction already appears in the binary environment that we introduced in Section 3.1 to which we now return. In fact, the following results are an immediate consequence of Proposition 1.

Proposition 5 (Data and Advertising)

1. *For all* $0 < c < 1$*, the monopoly price of a cookie is:*

$$p^*(c) = (1/2)\min\{c, 1-c\}.$$

2. *The equilibrium sales of cookies are given by the targeted set* $A(c, p^*(c))$*:*

$$A(c, p^*(c)) = \begin{cases} [0, c/2] & \text{if } \; c < 1/2; \\ [(1+c)/2, 1] & \text{if } \; c \geq 1/2. \end{cases}$$

3. *The equilibrium price, sales and profits of the data provider are single peaked in* c*.*

The data provider induces positive targeting when the cost of advertising c is sufficiently high and negative targeting when the cost of advertising is low. In consequence, the value of information is highest for intermediate levels of c. As a result, both the price of the data and the profits of the data provider are non-monotone in c. Intuitively, in the absence of information, advertisers choose either $q_0 = 0$ or $q_0 = 1$, depending on the cost of the advertising space c. For very low *and* very high values of c, the availability of data modifies the optimal action only on a limited set of consumers. Consequently, the willingness to pay for information is also limited.

The binary-action environment suggests which market conditions are more conducive to the profitability of a data provider. Perhaps contrary to a first intuition, niche markets with a high cost of advertising space and few profitable consumers are not necessarily the best environment. While the availability of data would have a large impact (demands for advertising would be nil without information), the data provider's profits are constrained by the low levels of surplus downstream. Instead, markets with relatively large fractions of both profitable and unprofitable consumers yield a higher value of information, which translates into higher prices for data and higher provider profits.

In the following subsections, we take the price of advertising as given, and we focus on the role of the market structure in the data sector. In Section 6, we take a more comprehensive look at the interplay of markets for data and advertising. In particular, we derive conditions under which a data provider and a publisher of advertising space have aligned or conflicting interests, and we leverage our results from Section 3 to explore how the data provider can profitably influence the price of advertising.

4.2 Data Sales Fragmentation

We have so far assumed a monopoly structure for the data industry. While the leading firms in this industry may currently hold considerable market power, the industry structure is evolving rapidly. Therefore, we assess the consequences of competition among sellers, and of the structure of the data industry in general. In particular, we focus on the externalities that each seller's price imposes on the other sellers through the composition of the advertisers' residual set.

Formally, we consider a continuum of data sellers, and we assume that each seller has exclusive information about one consumer segment i. Thus, each seller sets the price for one cookie only. This assumption corresponds literally to a market where individual users are able to sell their own data. It is also very closely related to the business model of the *data exchange*, where a data provider does not buy and resell information, but rather offers a platform for matching individual buyers and sellers, who set their own prices.[11]

For simplicity, we assume conditions under which positive targeting is optimal, and consider an advertiser's marginal willingness to pay as a function of the targeted set $A = [v_2, \bar{v}]$. The willingness to pay $p(v, v_2)$, for a cookie with valuation $v < v_2$ (the inverse demand

[11] For a description of integrated sales and separate billing, see *Cost vs. Value: Third-Party Targeting Data in the Demand-Side-Platform and Exchange Landscape*, Ad Exchanger, February 14th, 2011. http://www.adexchanger.com/agencies/cost-vs-value. For a more detailed description of the data exchange model, see http://www.bluekai.com/bluekai-exchange.php.

function) is given by the differential profit with respect to cookie v:

$$p(v, v_2) \triangleq (\pi(v, q^*(v)) - \pi(v, q^*(\mathbb{E}[v' \mid v' \leq v_2]))) . \tag{9}$$

We look for a symmetric pricing equilibrium in which advertisers choose positive targeting. We can think of each seller choosing a threshold v to maximize profits given the advertisers' purchasing strategy, and all other sellers' prices, which are summarized by the threshold v_2. Thus, a symmetric equilibrium cutoff solves the following problem:

$$v_2 = \arg\max_{v} [p(v, v_2)(1 - F(v))] .$$

The key difference with the monopoly problem lies in the residual advertising intensity $q^*(\mathbb{E}[v' \mid v' \leq v_2])$, which cannot be influenced by the price of any individual seller. More precisely, suppose the monopolist were to consider an expansion in the supply of cookies, and hence a lowering of the threshold v_2. By expanding the supply, he would reduce the gap between complete and prior information profits for the marginal consumer v_2. Naturally then, the monopolist would have to lower the price. But at the same time, the composition of the residual set will have changed. In fact, the average value on the residual set will have decreased, and thus the advertising level on the residual set will be lower. But this means that the value of information for the marginal consumer just below the targeted set has increased, and hence the marginal buyer just below the threshold will have a higher value of information. Now, this effect provides an additional incentive to lower prices and expand supply for the monopolistic data provider. But competing sellers do not internalize the positive externality present across cookie sales. Higher prices under fragmented data sales are then due to the lack of a strong compensating effect.

The fragmented data sales is illustrative of a more general result. Suppose we were to consider n symmetric data sellers, each holding information about a measure $1/n$ of consumers distributed identically according to $F(v)$. The n sellers set prices simultaneously. Consider now the trade-off facing a specific seller. She knows that, by lowering her price, all advertisers will purchase more from her, as well as from everyone else. This occurs because the action on the residual set will decrease, if only slightly. However, as for the case of fragmented sales, the compensation effect is attenuated in equilibrium by the fact that all other sellers are holding their prices fixed. Thus, in a symmetric equilibrium the price is increasing in n. As the number of sellers grows large, the equilibrium price approaches the price under fragmentation, where the action on the residual set is constant. When we contrast the equilibrium price with the case of a data monopolist, we obtain the following comparison.

Proposition 6 (Equilibrium under Data Fragmentation)

1. *The symmetric equilibrium price of cookies with a continuum of data sellers $\bar{p}$ is higher than the monopoly price p^*.*

2. *The symmetric equilibrium price with n independent and exclusive data sellers $p^*(n)$ is increasing in n, and approaches $\bar{p}$ as $n \to \infty$.*

Clearly, if many sellers would lead to duplication in the datasets, sellers would only be able to capture the incremental value of their information, thus driving prices down. In particular, there will exist a critical level of duplication for which the monopoly and the oligopoly prices are equal.

4.3 Reach of the Database

So far, we implicitly assumed that the monopolist's dataset covers all consumers, *i.e.,* that it has *maximal reach.* We now explore the implications of limited reach on the monopoly price of cookies, and on the equilibrium profits of the data provider and the advertisers.

We assume that the data provider owns information about a fraction $\beta < 1$ of all consumers. Advertisers know the distribution of match values of consumers present in the database, and of those outside of it. For simplicity, we further assume that the two distributions are identical, so that the measure of consumers in the dataset is given by $\beta F(v)$. Of course, in real-world data markets, consumers in a database may have different characteristics from those outside of it, and the presence of a cookie on a given consumer is informative *per se.*

Proposition 7 (Reach and Demand)
Assume exclusive (positive or negative) targeting is optimal. Then the advertisers' marginal willingness to pay $p(A, \beta)$ is increasing in β for all A.

Intuitively, the availability of more data improves the monopolist's profits. The more surprising part of the result is that demand for information shifts out as more consumers are reached by the database. The reason behind this result can be traced back to the effects of a larger database on the optimal action in the residual set $q^*(A^{\mathrm{C}})$. When positive targeting is optimal (so that $A = [v_2, \bar{v}]$), the average type in the residual set A^{C} is given by

$$\mathbb{E}[v \mid v \in A^{\mathrm{C}}] = \beta \mathbb{E}[v \mid v \leq v_2] + (1 - \beta) \mathbb{E}[v]. \tag{10}$$

Because the average type is decreasing in β for all A, the quantity of advertising demanded on the residual set is decreasing in β. Thus, the willingness to pay for information on the *marginal* consumer v_2 increases. A similar argument applies to the case of negative targeting.

This result does not, however, imply that the monopolist wishes to raise prices as the reach β increases. On the contrary, Proposition 8 identifies sufficient conditions under which a database with a wider reach is sold at a lower unit price.

Proposition 8 (Reach and Monopoly Price)
Suppose the matching costs are quadratic and values v are distributed according to $F(v) = v^{\lambda}$, with $v \in [0,1]$ and $\lambda \in (0,1)$. If exclusive targeting is optimal, then the monopoly price $p^*(\beta)$ is strictly decreasing in the reach β.

As the reach of the database β increases, the optimal monopoly price is pushed lower by two effects. First, the willingness to pay for any targeted set increases (Proposition 7), which makes raising price and restricting supply more costly. Second, the optimal action in the residual set is now more sensitive to the price of cookies. This is due to the compensating effect: the average consumer outside the targeted set becomes less likely to have a high match value; as a consequence, the quantity of advertising demanded on the residual set decreases faster as the targeted set expands. Both these effects induce the monopolist to lower price and expand supply as the database becomes less limited.

Two final remarks are in order. First, a reduction in price implies an increase in the *range* of data sold by the monopolist $[v_2, \bar{v}]$ as the reach β increases. Therefore, an increase in the reach β leads to higher data sales. Thus advertisers pay a lower price and access more information, which implies that their profits increase. This means that an increase in data availability can induce a Pareto improvement in the market for information.[12]

Second, note that the price of information is not necessarily continuous or monotone in the reach parameter β. In particular, jumps may occur when the targeting policy induced by the monopolist switches from joint (both positive and negative) targeting for low reach values β to exclusive (positive *or* negative) targeting for high reach values β.

5 Beyond Linear Pricing

We have focused so far on a fairly specific set of information structures (cookies-based) and pricing mechanisms (linear prices). We now generalize our analysis of data sales to address two closely related questions: *(i)* What is the optimal mechanism to sell information in

[12] In Section 6, we address the effect of higher data sales on the market for advertising.

our environment? *(ii)* Are there conditions under which pricing of individual cookies can implement the optimal mechanism?

Up to now, we assumed that the advertisers are symmetric in the distribution of the match values. Moreover, the advertisers attached the same willingness to pay to a consumer with match value v. Thus, from an ex-ante point of view, the advertisers are all identical, and their common ex ante value of information is assumed to be known to all market participants, including the data provider. Therefore, it is as if the data provider has complete information about the preferences of the advertisers. Now, in this precise setting, a data provider who could choose among *unrestricted* information and pricing policies would be able to extract the entire ex ante surplus from the advertisers. In particular, he could charge a bundle price for the entire database equal to the ex-ante value of information.

However, the assumption of complete information appears to be too strong in the world of "big data" where advertisers are heterogeneous. In this section, we shall therefore allow for a private-information component in the advertisers' willingness to pay to match with a consumer with characteristics v. Thus, we consider advertisers who differ in their marginal willingness to pay, denoted by $\theta \in \Theta = [0, 1]$. Extending the earlier expression (1), the net value of a match is now given by:

$$\pi (v, q, \theta) \triangleq \theta v q - c \cdot m (q) .$$

The marginal willingness to pay θ is private information to each advertiser and is distributed in the population of advertisers according to a continuous distribution function $G (\theta)$ with density $g (\theta)$. For this section, we return to the binary decision environment of Section 3.1, and restrict attention to binary decisions $q \in \{0, 1\}$ of the advertiser (or alternatively linear matching cost $m (q) = q$). The net value of a match is then given by, extending the earlier expression (4):

$$\pi^* (v, \theta) \triangleq \max \{\theta v - c, 0\} .$$

Thus, for advertising to generate positive value, the realization of θ must exceed c as $v \in [0, 1]$.

We now explore the data provider's ability to screen advertisers by offering different information policies, and by pricing the amount of information in a nonlinear way. We begin our analysis with *noiseless* information policies and characterize the optimal mechanism within the class of noiseless information structures. In fact, we can establish that noiseless information policies remain optimal even when we consider arbitrary information structures. This result requires substantial additional language and notation for general information structures that go beyond those used in the main text, and is thus relegated to the Appendix B.

With binary actions, the socially efficient information policy can be induced by a threshold $v^*(\theta)$ that informs advertisers perfectly and without noise about the match value v if and only if v exceeds the threshold $v^*(\theta)$ given by:

$$v^*(\theta) = \frac{c}{\theta}. \tag{11}$$

In other words, the data provider can attain the efficient allocation of information through an information policy based on cookies. Under the efficient information policy, each advertiser receives information about every realization of v such that $v \geq v^*(\theta)$. Each advertiser would then adopt positive targeting, *i.e.*, contact all consumers it receives a signal about, and ignore the residual users. The expected gross value of the efficient information policy for an advertiser with willingness to pay θ is:

$$w^*(\theta) \triangleq \int_{\frac{c}{\theta}}^{1} (\theta v - c)\,\mathrm{d}F(v).$$

Now consider an arbitrary noiseless information policy with threshold x. The value of this information structure to an advertiser with willingness to pay θ is given by:

$$w(\theta, x) \triangleq \int_{x}^{1} (\theta v - c)\,\mathrm{d}F(v). \tag{12}$$

Note the submodularity property of $w(\theta, x)$, namely that $\partial^2 w(\theta, x)/\partial\theta\partial x = -v < 0$. Therefore, any implementable information policy leads to more data, and hence lower thresholds x, being assigned to advertisers with higher willingness to pay θ. Given the noiseless nature of the information policy, the above problem (12) is akin to a nonlinear pricing problem, where the quantity variable is the amount of information, or the number of cookies sold.

In the associated direct revelation mechanism, each advertiser communicates his willingness to pay, and in exchange is offered a set of cookies and a price for the bundle of cookies. The set of cookies is determined by the threshold $v^*(\theta)$ and hence the associated quantity of cookies is

$$Q(\theta) \triangleq 1 - F(v^*(\theta)),$$

and we denote the transfer payment in the direct mechanism by $T(\theta)$. As in the standard analysis of revenue-maximizing mechanisms, we impose a regularity condition such that the local incentive conditions generate the requisite monotone allocation, which in this context is simply the requirement that the "virtual utility"

$$\theta - \frac{1 - G(\theta)}{g(\theta)} \tag{13}$$

is increasing in θ. We maintain this restriction for the remainder of this section.

Proposition 9 (Information Policy)
The optimal information policy involves a noiseless information policy with threshold $v^*(\theta)$ given by:

$$v^*(\theta) = \frac{c}{\theta - \frac{1-G(\theta)}{g(\theta)}}. \tag{14}$$

Perhaps the surprising element in the determination of the information policy is that the distributional information about the match values (*i.e.*, $f(v)$ or $F(v)$) does not appear in the description of the optimal information policy. This results from the additivity of the utility of all types θ in the number of user contacts, *i.e.*, differences in willingness to pay originate from the match values v only.

The direct mechanism establishes some key properties of the information policy. In particular, $T(\theta)$ and $Q(\theta)$ are strictly increasing in θ, as shown in Proposition 10.1. But a related, indirect mechanism speaks more directly to the problem of data selling and access to the database. Namely, the data provider could specify a nonlinear pricing scheme, or conversely a price for incremental access to the database. With $Q(\theta)$ strictly increasing in θ, we can define a nonlinear pricing scheme, which associates every quantity Q with the transfer of the corresponding type $Q^{-1}(\theta)$:

$$P(Q) \triangleq T\left(Q^{-1}(\theta)\right).$$

We define the price $p(Q)$ as the price for incremental access to the database, or the marginal price that we can readily interpret as the price of an additional cookie:

$$p(Q) \triangleq P'(Q).$$

We can then establish, under slightly stronger regularity conditions than (13), that the incremental pricing $p(Q)$ implements the direct mechanism as an indirect mechanism. In fact, the data provider offer access to additional cookies at a declining price that mirrors the logic of quantity discounts as in Maskin and Riley (1984).

Proposition 10 (Prices and Quantities)

1. *The number of cookies sold,* $Q(\theta)$ *and the transfer* $T(\theta)$ *are increasing in* θ.

2. *The incremental cookie price* $p(Q)$ *is decreasing in* Q *and decentralizes the direct optimal mechanism if* $(1 - G(\theta))/g(\theta)$ *is decreasing.*

Thus, the data provider can decentralize the optimal direct mechanism by allowing advertisers to access a given portion of the database, with volume discounts for those who demand a larger amount of cookies. This establishes an equivalent implementation of the optimal mechanism, based on advertiser self-selection of a subset of cookies. In this sense, we can view linear prices as simple approximations to the optimal mechanism in this particular case.[13]

6 Interaction between Data and Advertising Markets

We finally return to linear pricing of cookies in order to examine the interaction between the markets for data and online advertising. We seek to assess (a) the effect of the price of data p on the advertising publisher's revenue, and (b) the effect of the unit price of advertising c on the data provider's revenue. Understanding these cross-market externalities will allow us to study the data provider's pricing problem in richer scenarios. In particular, in Section 6.2, we consider the monopoly price of cookies when the unit price of advertising is determined endogenously; and in Section 6.3, we let the data provider sell cookies either to the advertisers or to the publisher, and we determine which side of the advertising market yields higher profits.[14]

6.1 Demand for Advertising

The effect of the price of information on the total demand for advertising space and the impact of the cost of advertising on the demand for data are unclear *a priori*. For instance, the total demand for advertising space may increase or decrease in the amount of information available to advertisers, depending on whether the data is used for positive or negative targeting. Likewise, a lower marginal cost of advertising space increases the advertisers'

[13] See Rogerson (2003) for bounds on the loss in profits from simpler mechanisms such as linear pricing.

[14] In the online advertising industry, sellers often purchase third-party data. For example, The Economist acquires Bizo Private Audience Targeting through the BlueKai Data Exchange, in order to enable demographic targeting (http://www.economist.com/cookies-info). In Australia, Yahoo! uses Acxiom's offline data to improve its online targeting ability (http://www.acxiom.com/press-releases/2012/acxiom-announces-first-quarter-results/).

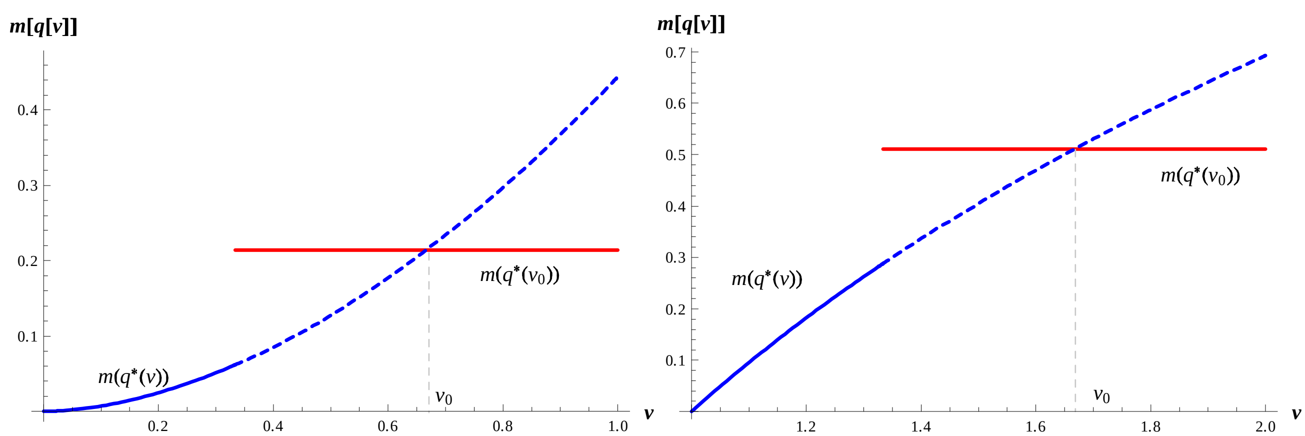
downstream surplus, but it also decreases the value of information by reducing the cost of advertising to the residual set.

To formalize these trade-offs, consider the total demand for advertising space as a function of the targeted set $A(c,p)$. Because any advertiser who wishes to generate match intensity q with a consumer must purchase an amount of space equal to $m(q)$, the total demand for advertising is given by

$$M(A) \triangleq \int_A m(q^*(v))\,\mathrm{d}F(v) + \int_{A^{\mathrm{C}}} m(q^*(A^{\mathrm{C}}))\mathrm{d}F(v). \tag{15}$$

We are interested in the effect of the amount of data sold $\mu(A)$ on the total demand for advertising $M(A)$. Suppose for now that negative targeting is optimal, *i.e.*, the residual set is given by $A^{\mathrm{C}}(c,p) = [v_1, \bar{v}]$ for some threshold $v_1(c,p) > \underline{v}$. As the price of data increases, v_1 decreases and the publisher replaces $m(q^*(v_1))$ with $m(q^*([v_1,\bar{v}]))$, which is higher. At the same time, the average match value $\mathbb{E}[v \mid v \geq v_1]$ decreases, thus reducing the match intensity with every consumer in the residual set. Figure 5 compares the demand for advertising $m(q(v))$ for fixed targeted and residual sets, under two different matching cost functions.

Figure 5: Total Demand for Advertising



$m(q) = q^b/b,\ b = 9/4$ $\qquad$ $m(q) = -\ln(1-q)$

We again relate the determinants of the demand for information to the properties of the complete-information demand for advertising. The total demand for advertising (*i.e.*, the area under the solid lines in Figure 5) is increasing in the measure of the targeted set A when the complete-information demand for advertising is convex in v. Figure 5 therefore helps to clarify the effect of the marginal cost of advertising c on the data provider's revenue. Suppose the unit cost of advertising space c increases. Everything else constant, this induces advertisers to reduce their demand for advertising space. In particular, if the complete-information demand for advertising is convex in v, a reduction in the amount of data purchased reduces the total advertising expenditure. Thus, when c increases, advertisers' marginal willingness to pay for advertising decreases and so does the revenue of the data provider.

Proposition 11 formalizes the interaction of the data and advertising markets by relating the nature of the cross-market externalities to the properties of the matching cost function.

Proposition 11 (Market Interaction)
Assume exclusive (positive or negative) targeting is optimal.

1. *If $m'(q)$ is log-concave, then the data provider's revenue is decreasing in c, and the publisher's revenue is decreasing in p.*

2. *If $m'(q)$ is log-convex, then the data provider's revenue is increasing in c, and the publisher's revenue is increasing in p.*

In the proof of Proposition 11, we show that convexity of the complete-information demand for advertising is equivalent, in terms of the primitives of our model, to the log-concavity of the marginal cost of matching. In turn, log-concavity and log-convexity of $m'(q)$ are sufficient to establish the sign of the cross-market externality. Finally, the conditions in Proposition 11 are related to the optimality of positive vs. negative targeting (see Proposition 4). In particular, positive targeting requires convexity of $q^*(v)$, while negative cross-market externalities require convexity of the composite function $m^*(q(v))$. Thus, the optimality of positive targeting implies negative cross-market externalities, but not vice-versa.

6.2 Endogenous Cost of Advertising

We now leverage our results on market interaction to assess how the monopoly price of cookies responds to competition in the "downstream" market for advertising. We introduce a fixed supply of space M for each user i. This may correspond to a limit on the actual physical space on web pages that the user can access, or to a limit on the user's attention span.

We consider a game in which the data provider sets the price of cookies, advertisers buy information, and then compete for a fixed supply of advertising space. In equilibrium, the prices of cookies and advertising space are such the data provider and the advertisers maximize profits, and the advertising market clears. In other words, the equilibrium price of advertising $c^*(p, M)$ satisfies the following market-clearing condition,

$$M = \int_{A(c,p)} m\left(q^*(v)\right) \mathrm{d}F(v) + \int_{A^{\mathrm{C}}(c,p)} m(q^*(A^{\mathrm{C}}(c,p)))\mathrm{d}F(v)$$

By controlling the price of cookies p, the data provider can profitably influence the total demand for advertising, thus affecting the equilibrium price of advertising and hence the demand for cookies. We now apply our earlier results, and compare the monopoly price of cookies under exogenous and endogenous prices of advertising space.

Consider the case of convex complete-information demand for advertising. In this case, the price of advertising imposes a negative externality on the data provider. Thus, compared to the case of an exogenous c, the data provider wishes to reduce congestion downstream in order to keep the equilibrium price c^* low. Because demand for advertising space is decreasing in p, the data provider must raise its price.

Now consider the opposite case of concave complete-information demands. We know the data provider wishes to keep c^* high in order to increase the demand for information. However, the total demand for advertising is now increasing in p. Again, the monopolist data provider wishes to raise the price of data, compared to the case of exogenous cost of advertising.

To summarize, while the cross-market payoff externalities depend on the matching cost function, the strategic implications of downstream competition for the data provider are more clear: under the sufficient conditions of Proposition 11, the data provider profitably increases the price of cookies in response to fiercer competition for advertising space. In turn, an increase in the price of data may benefit the publisher of the advertising space. This occurs when returns to advertising decrease sufficiently fast that the complete-information advertising demand is concave in v. We now explore the role of the matching technology when we allow the data provider to choose between selling information to the advertisers or to the publisher.

6.3 Selling Cookies to the Publisher

The previous discussion highlighted the possibility that online publishers benefit from the direct sale of information to the advertisers. *A fortiori*, the publisher may benefit from the indirect sale of information, *i.e.*, from purchasing cookies and releasing match-value information to the advertisers. Proposition 11 identified conditions under which the demand of advertising space is increasing in the amount of data sold $\mu(A)$.

We now restrict attention to technologies for which the publisher has a positive value of information ($i.e.$, we assume $m\left(q^{*}\left(v\right)\right)$ is convex). We characterize the price of information that the data provider can charge to either side of the advertising market. In our framework, all cookies are symmetric from the point of view of the publisher, and its revenue is linear in the measure of cookies bought. Therefore, the publisher's marginal willingness to pay for data is constant, and it is given by

$$p = c \cdot M\left(V\right),$$

where $V = [\underline{v}, \bar{v}]$ and the total advertising demand $M\left(\cdot\right)$ is defined in (15). Thus, the data provider faces a flat demand curve when selling to the publisher, and a downward-sloping demand curve when selling to the advertisers. In Figure 6, we compare the data provider's profits as a function of the "target buyers" as we vary the matching cost function. In particular, we consider uniformly distributed match values and power cost functions $m\left(q\right) = q^{b}/b$. In this example, as b increases, the marginal returns to advertising decline faster.

Figure 6: Selling to Either Side

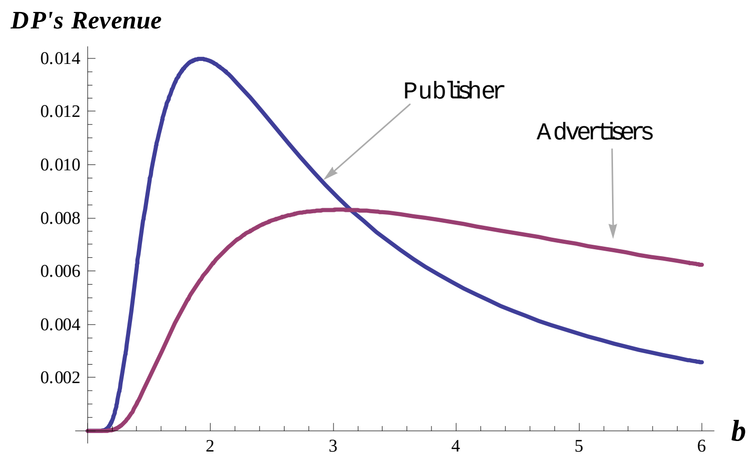


Figure 6 then suggests that the publishers have a higher willingness to pay for data when the marginal returns to advertising are fairly high. Conversely, if the decline in the marginal returns is steep (so that complete-information demands are nearly linear in v), the publisher's willingness to pay is low, and the data provider prefers selling information to the advertisers.

7 Concluding Remarks

In this paper, we have explored the sale of individual-level information in a setting that captures the key economic features of the market for third-party data. Specifically, in our model, a monopolistic data provider determines the price to access informative signals about each consumer's preferences.

Our first set of results characterized the demand for such signals by advertisers who wish to tailor their spending to the match value with each consumer. We showed how properties of the complete-information profit function determine the optimality of an information-purchasing strategy that achieves positive targeting, negative targeting, or both. Turning to monopoly pricing of cookies, we established that the ability to influence the composition of the advertisers' targeted and residual sets was the key driver of the optimal (linear) prices. As a consequence, both the reach of the monopolist's database and the concentration of data sales provide incentives to lower prices.

We considered an environment in which advertisers differ in their willingness to pay, and we showed that cookies-based pricing can be part of an (approximate) optimal mechanism for the sale of information. In particular, we showed that the data provider can decentralize the optimal mechanism by offering a nonlinear pricing schedule for cookies.

Finally, we explored the interaction between the markets for data and advertising. We obtained conditions under which the price in each market has a negative impact on the seller's profits in the other market, and we showed that the publisher of advertising can, but need not, benefit from large data sales.

We, arguably, made progress towards understanding basic aspects of data pricing and data markets. We did so by making a number of simplifying assumptions. A more comprehensive view of data markets would require a richer environment. In the present model, neither the advertiser nor the publisher had access to any information about the consumers. In reality, advertisers and (more prominently) large publishers and advertising exchanges maintain databases of their own. Thus, the nature of the information sold, and the power to set prices depends on the allocation of information across market participants. An interesting question in this context is related to the effects of privacy regulation (*e.g.*, banning the sale of information) on the allocative effects of information markets.

Moreover, online data transactions are inherently two-sided. Presently, we analyzed the price charged by the data provider to the advertisers. But there are cost of acquiring the data, either from individuals, publisher or advertisers. Ultimately, the cost of acquiring information for the data provider should be related to the value of privacy, which might limit the availability of data, or at least raise its price.

Appendix A

Proof of Proposition 1. Suppose the advertisers' optimal action on the residual set is given by $q^*(A^C) = 0$. The value of the marginal cookie is then given by $\max\{0, v - c\}$, which is increasing in v. We show that the value of information is strictly monotone in v. Notice that adding higher-v cookies to the targeted set does not change the optimal action on the residual set, because it lowers the expected value of a consumer $v \in A^C$. Thus, if advertisers buy cookie v, they also buy all cookies $v' > v$. Conversely, if the optimal action on the residual set is given by $q^*(A^C) = 1$, the value of the marginal cookie is $\max\{0, c - v\}$. By a similar argument, the value of information is strictly decreasing in v: if advertisers buy cookie v, they also buy all cookies $v' < v$.

Now consider the advertiser's profits under positive and negative targeting. In the former case, the advertisers' profits are given by

$$\pi_+(c,p) \triangleq \max_{v} \int_{v}^{1} (v - c - p)\,\mathrm{d}F(v) = \int_{c+p}^{1} (v - c - p)\,\mathrm{d}F(v).$$

In the latter case, profits are given by

$$\pi_-(c,p) \triangleq \max_{v} \left[\int_{v}^{1} (v - c)\,\mathrm{d}F(v) - pF(v)\right] = \int_{c-p}^{1} (v - c)\,\mathrm{d}F(v) - pF(c-p).$$

Now consider the difference

$$\pi_+(c,p) - \pi_-(c,p) = p\left(F(c-p) + F(c+p) - 1\right) - \int_{c-p}^{c+p} (v - c)\,\mathrm{d}F(v). \qquad (16)$$

Under the uniform distribution, the second term in (16) is nil, while the first is equal to $p(2c-1)$, which establishes the result. ■

Proof of Proposition 2. Suppose towards a contradiction that the optimal residual set A^C is not an interval. Let $q_0 = q^*(A^C)$ denote the match intensity with all consumers in the residual set. By equation (6), we know q_0 is the optimal match intensity for the average type $\bar{v}_A = \mathbb{E}[v \mid v \notin A]$. Suppose $\bar{v}_A \in A$. Now consider two consumers with $v'' > v'$ and $q^*(v'') > q^*(v') > q_0$ such that the firm buys cookie v' but not v''. If A^C is not an interval, either such a pair exists, or there exists a pair with $v'' < v'$ and $q^*(v'') < q^*(v') < q_0$ such that the firm buys cookie v' but not v''. Consider the former case, and compute the change in profits obtained by swapping cookies, *i.e.*, purchasing (an equal number of) cookies v'' instead of cookies v'. Define the difference between complete and incomplete-information profits as

$$\Delta(v, q_0) = v(q^*(v) - q_0) - c(m(q^*(v)) - m(q_0)),$$

and notice that $\Delta_v(v, q_0) = (q^*(v) - q_0)$. Therefore $q^*(v'') > q^*(v') > q_0$ implies $\Delta(v'', q_0) > \Delta(v', q_0)$. Because the advertiser gains $\Delta(v'', q_0)$ and loses $\Delta(v', q_0)$, it follows that the swap strictly improves profits. An identical argument applies to the case of $q^*(v'') < q^*(v') < q_0$. Finally, if $\bar{v}_A \notin A$, then a profitable deviation consists of not purchasing $\bar{v}_A$: advertisers avoid paying a positive price, and the optimal action on the residual set does not change. ■

Proof of Proposition 3. If costs are quadratic, so are the complete-information profits. By symmetry of the distribution, $v_0 = \mathbb{E}[v \mid v \in [v_0 - \varepsilon, v_0 + \varepsilon]]$ for any $\varepsilon > 0$. The marginal value of information is then given by

$$p(v) = \pi^*(v) - (vq^*(v_0) - cm(q^*(v_0))) = (v_0 - v)^2 / 4c.$$

Solving for v_0 yields the optimal residual set as a function of p and c. ■

Proof of Proposition 4. Consider the necessary conditions for the optimal residual set A^{C} to be given by an interior interval $[v_1, v_2]$. Denote the expected value of v on the residual set by

$$v_0 \triangleq \mathbb{E}[v \mid v \in A^{\mathrm{C}}].$$

It follows that $q_0 \triangleq q^*(A^{\mathrm{C}}) = q^*(v_0)$, and by the envelope theorem $q_0 = \pi'(v_0)$. The marginal value of information at v is then given by $\pi(v) - (\pi(v_0) + (v - v_0)\pi'(v_0))$, and its derivative with respect to v is given by $\pi'(v) - \pi'(v_0)$. Optimality of an interior residual set requires that the marginal value of information is equal to p at the two extremes *i.e.*,

$$\int_{v_1}^{v_2} (\pi'(v) - \pi'(v_0))\,\mathrm{d}v = 0.$$

Under concavity of $\pi'(v)$, however, we have

$$\int_{v_1}^{v_2} (\pi'(v) - \pi'(v_0))\,\mathrm{d}v \leq \int_{v_1}^{v_2} \pi''(v_0)(v - v_0)\,\mathrm{d}v,$$

which is nonpositive if $f(v)$ is nondecreasing. This implies negative targeting. A similar last step implies positive targeting.

Finally, we relate the curvature of the profit function to that of the match cost function. The envelope theorem implies $\pi'(v) = q^*(v)$, and implicit differentiation of the first order condition yields

$$\pi''(v) = \frac{1}{cm''(q^*(v))}.$$

Because $q^*(v)$ is strictly increasing, we conclude that $\pi'''(v) > 0$ if and only if $m'''(q) < 0$.■

Proof of Proposition 5. (1.) We know from Proposition 1 that advertisers choose the following targeted set:

$$A(c,p) = \begin{cases} [0, \max\{c-p, 0\}] & \text{if} \quad c < 1/2; \\ [\min\{c+p, 1\}, 1] & \text{if} \quad c \geq 1/2. \end{cases} \tag{17}$$

Thus, under the uniform distribution, the monopoly price of cookies is given by

$$p^*(c) = \begin{cases} \arg\max_p [p(c-p)] & \text{if} \quad c < 1/2, \\ \arg\max_p [p(1-c-p)] & \text{if} \quad c \geq 1/2, \end{cases}$$

and therefore $p^*(c) = (1/2)\min\{c, 1-c\}$.
(2.) It follows from (17) that $A(c, p^*(c)) = [0, c/2]$ if $c < 1/2$ and $A(c, p^*(c)) = [(1-c)/2, 1]$ if $c \geq 1/2$.
(3.) The single-peakedness of prices $p^*(c)$, sales $\mu(A(c, p^*(c)))$, and hence profits, is immediate from parts (1.) and (2.). ■

Proof of Proposition 6. (1.) Let

$$p(v,x) = \pi(v) - \pi(v, q^*(\mathbb{E}[v < x])).$$

Under monopoly, the data provider's chooses the marginal cookie v_2 to solve the following problem:

$$\max_v [p(v,v)(1-F(v))].$$

The optimal v_2^* is then given by the solution v to the following first-order condition:

$$-p(v,v)f(v) + \partial p(v,v)/\partial v + \partial p(v,v)/\partial x = 0.$$

Conversely, in the symmetric equilibrium with a continuum of sellers, the equilibrium marginal cookie $\bar{v}_2$ is given by the solution v to the following condition

$$-p(v,v)f(v) + \partial p(v,v)/\partial v = 0.$$

However,
$$\frac{\partial p\left(v, v_2^*\right)}{\partial x}=-\frac{\partial \pi\left(v, q^*\left(\mathbb{E}\left[v<x\right]\right)\right)}{\partial q} \frac{\partial q^*}{\partial v} \frac{\partial \mathbb{E}\left[v<x\right]}{\partial x}<0,$$
because $q^*(v)$ is strictly increasing in v, and therefore $\partial \pi(v, q) / \partial q>0$ for all $q<q^*(v)$. Therefore, the price under competition $\bar{p} \triangleq p\left(\bar{v}_2, \bar{v}_2\right)$ is higher than the monopoly price $p^* \triangleq p\left(v_2^*, v_2^*\right)$.

(2.) We look for a symmetric equilibrium in the price-setting game with n data providers. Let $p_j=p_2$ for all $j \neq 1$ and characterize the advertisers' demand as a function of $\left(p_1, p_2\right)$. If positive targeting is optimal, advertisers buy cookies $v \in\left[v_1, \bar{v}\right]$ from seller $j=1$ and $v \in\left[v_2, \bar{v}\right]$ from sellers $j \neq 1$. In particular, the thresholds $\left(v_1, v_2\right)$ satisfy the following equations:
$$\begin{aligned}
\pi\left(v_1\right)-\pi\left(v_1, q^*(\hat{v})\right) & =p_1 \\
\pi\left(v_2\right)-\pi\left(v_2, q^*(\hat{v})\right) & =p_2,
\end{aligned}$$
where
$$\hat{v}\left(p_1, p_2\right)=\frac{\mathbb{E}\left[v \mid v<v_1\right]+(n-1) \mathbb{E}\left[v \mid v<v_2\right]}{n}.$$
Note that $p_1>p_2$ implies $v_1>v_2$. Now rewrite the profit function of seller $j=1$ as
$$\Pi_1=\left(\pi\left(v_1\right)-\pi\left(v_1, q^*(\hat{v})\right)\right)\left(1-F\left(v_1\right)\right).$$
At a symmetric equilibrium where $v_j \equiv v$, the first-order condition of seller 1 is given by
$$\left(\pi\left(v_1\right)-\pi\left(v_1, q^*(\hat{v})\right)\right) \frac{f\left(v_1\right)}{1-F\left(v_1\right)}=\pi^{\prime}\left(v_1\right)-\frac{\partial \pi\left(v_1, q^*(v)\right)}{\partial v_1}-\frac{\partial \hat{v}}{\partial v_1} \frac{\mathrm{d} q^*(\hat{v})}{\mathrm{d} \hat{v}} \frac{\partial \pi\left(v_1, q^*(v)\right)}{\partial q}.$$
Notice that both $d q^*(\hat{v}) / d \hat{v}$ and $\partial \pi\left(v_1, q^*(v)\right) / \partial q$ on the right-hand side are positive. Therefore, because
$$\frac{\partial \bar{v}}{\partial v_1}=\frac{1}{n} \frac{\partial \mathbb{E}\left[v \mid v<v_1\right]}{\partial v}$$
is decreasing in n, the symmetric equilibrium threshold $v^*(n)$ is increasing in n, and so is the price $p^*(n)$. ■

Proof of Proposition 7. Under positive targeting, the marginal willingness to pay $p(v, \beta)$ for a targeted set $A=[v, \bar{v}]$ is given by
$$p(v, \beta) \triangleq \pi^*(v)-\pi\left(v, q_0(v, \beta)\right),$$

where

$$q_0(v,\beta) \triangleq q^*\left(\beta \mathbb{E}_F[v' \mid v' < v] + (1-\beta)\mathbb{E}_F[v']\right).$$

The derivative of the inverse demand function with respect to the reach β is given by

$$\frac{\partial p(v,\beta)}{\partial \beta} = -\left(v - cm'(q_0(v,\beta))\right) q^{*\prime}(\cdot)\left(\mathbb{E}_F[v' \mid v' < v] - \mathbb{E}_F[v']\right). \tag{18}$$

The first two terms in (18) are positive: profits $\pi(v, q_0)$ are increasing in q because $q_0(v,\beta) < q^*(v)$; the complete information quantity $q^*(\cdot)$ is strictly increasing; and difference of the conditional and unconditional expected values is negative. Therefore, the marginal willingness to pay $p(v,\beta)$ is increasing in β. ■

Proof of Proposition 8. Under the quadratic matching costs and distributional assumption ($F(v) = v^\lambda$, $\lambda < 1$), the inverse demand $p(v,\beta)$ for a targeted set $A = [v, \bar{v}]$ can be written as

$$p(v,\beta) = \frac{1}{2c}\left(\frac{1-\beta-v(1-\beta+\lambda)}{1+\lambda}\right)^2.$$

Therefore, the monopolist maximizes

$$\Pi(v,\beta) \triangleq \left(1-v^\lambda\right) p(v,\beta).$$

The first-order condition $\partial\Pi/\partial v = 0$ can be solved for the inverse function $\beta^*(v)$, which is given by

$$\beta^*(v) = \frac{(1+\lambda)\left(2 - v^\lambda(2+\lambda)\right) + v^{\lambda-1}\lambda}{2\left(1-v^\lambda\right) + \lambda(1-v)v^{\lambda-1}}.$$

Differentiating with respect to v, one obtains that $\beta^*(v)$ is decreasing in v if

$$v > \hat{v}(\lambda) \triangleq \left(\frac{2(1-\lambda)}{2+\lambda}\right)^{\frac{1}{\lambda}}.$$

Substituting $v = \hat{v}(\lambda)$ into $\beta^*(v)$ one obtains $\beta^*(\hat{v}(\lambda)) > 1$ for all $\lambda \in (0,1)$. Thus, $v^*(\beta,\lambda) > \hat{v}(\lambda)$, which in turn implies $\beta^*(v)$ is decreasing. Furthermore, substituting $\beta^*(v)$ into the inverse demand, we obtain the monopoly price as a function of the optimal range v,

$$p(v,\beta^*(v)) = 2\frac{\lambda^2 v^2\left(1-v^\lambda\right)^2}{(1+\lambda)^2\left(-2v + v^\lambda(2v-(1-v)\lambda)\right)^2}.$$

Finally, one can show that the sign of the total derivative $dp(v,\beta^*(v))/dv$ depends on the sign of $-1+\lambda-\lambda v+v^\lambda$, which is positive for all $(\lambda, v) \in [0,1]^2$. Therefore, the optimal range v is decreasing in β and so is the monopoly price. ■

Proof of Proposition 9. Starting with the value of information $w(\theta, x)$ given in (12), we have

$$\frac{\partial w(\theta, x)}{\partial x} = -(\theta x - c) f(x), \tag{19}$$

with

$$\frac{\partial w(\theta, x)}{\partial \theta} = \int_x^1 v \mathrm{d}F(v),$$

and hence

$$\frac{\partial^2 w(\theta, x)}{\partial \theta \partial x} = -x f(x). \tag{20}$$

Because of this submodularity property, higher types θ should receive lower cutoffs x.

The optimal information allocation and pricing can then be solved via the virtual utility, and is given by:

$$\frac{\partial w(\theta, x)}{\partial x} = \frac{1 - G(\theta)}{g(\theta)} \frac{\partial^2 w(\theta, x)}{\partial x \partial \theta},$$

and after using (19) and (20) we obtain the result in (14). ■

Proof of Proposition 10. (1.) It follows immediately from the derivation of the optimal threshold (14) that the quantity of cookies sold, $Q(\theta)$, is increasing, given that the virtual utility $\theta - (1 - G(\theta))/g(\theta)$ is increasing in θ. We can derive the associated transfer rule from the gross utility of the buyer:

$$w(\theta, x(\theta)) = \theta \int_{x(\theta)}^1 v \mathrm{d}F(v) - c \int_{x(\theta)}^1 \mathrm{d}F(v). \tag{21}$$

After all, the indirect utility in any incentive compatible mechanisms is given by standard Mirrlees formula as the integral over the local incentive constraints (and information rents) :

$$W(\theta) \triangleq \int_{\underline{\theta}}^{\theta} \frac{\partial w(\theta', x(\theta'))}{\partial \theta'} \mathrm{d}\theta' = \int_{\underline{\theta}}^{\theta} \int_{x(\theta')}^1 v dF(v) \mathrm{d}\theta'. \tag{22}$$

The associated transfers (in the direct mechanism) are then given by

$$T(\theta) = w(\theta, x(\theta)) - W(\theta).$$

Thus the transfer payment is given by

$$T(\theta) = -c \int_{x(\theta)}^1 g(v) \mathrm{d}v - \int_{\underline{\theta}}^{\theta} \theta' x(\theta') g(x(\theta')) \frac{\mathrm{d}x(\theta')}{\mathrm{d}\theta'} \mathrm{d}\theta', \tag{23}$$

and differentiating (23) with respect to θ we find:

$$\begin{aligned} T'(\theta) &= cf(x(\theta))\frac{\mathrm{d}x(\theta)}{\mathrm{d}\theta} - \theta x(\theta) f(x(\theta))\frac{\mathrm{d}x(\theta)}{\mathrm{d}\theta}\mathrm{d}\theta \\ &= -\frac{\mathrm{d}x(\theta)}{\mathrm{d}\theta} f(x(\theta))(\theta x(\theta) - c) \geq 0, \end{aligned} \tag{24}$$

where the inequality follows from (14) and the monotone virtual utility assumption.
(2.) We can rewrite the transfer also in terms of the threshold $x(\theta)$ or the quantity sold $Q(\theta) = 1 - F(x(\theta))$, and hence $P(Q(\theta))$, and so using (24), we get

$$t'(x)\frac{\mathrm{d}x(\theta)}{\mathrm{d}\theta} = -\frac{\mathrm{d}x(\theta)}{\mathrm{d}\theta} f(x(\theta))(\theta x(\theta) - c) \Leftrightarrow t'(x) = -f(x(\theta))(\theta x(\theta) - c).$$

Now, the unit price per cookie sold at realization $x(\theta)$ is given by:

$$\frac{t'(x(\theta))}{f(x(\theta))} = -(\theta x(\theta) - c),$$

and using the solution $x(\theta) = v^*(\theta)$ from (14), we get

$$\frac{t'(x(\theta))}{f(x(\theta))} = -\left(\theta\frac{c}{\theta - \frac{1-G(\theta)}{g(\theta)}} - c\right) = -\frac{c\frac{1-G(\theta)}{g(\theta)}}{\theta - \frac{1-G(\theta)}{g(\theta)}}.$$

Thus if $\theta' > \theta$ and hence $x(\theta') < x(\theta)$, then

$$\frac{t'(x(\theta'))}{f(x(\theta'))} < \frac{t'(x(\theta))}{f(x(\theta))},$$

and thus the price per cookie is decreasing. ■

Proof of Proposition 11. We first establish a property of the complete-information demands for advertising. Differentiating $m(q^*(v))$ with respect to v, we obtain

$$\frac{\mathrm{d}m(q^*(v))}{\mathrm{d}v} = m'(q^*(v))\frac{\mathrm{d}q^*(v)}{\mathrm{d}v} = \frac{m'(q^*(v))}{cm''(q^*(v))}.$$

Therefore, the demand for advertising space is convex in v if $m''(q)/m'(q)$ is decreasing in q, *i.e.*, $m'(q)$ is log-concave. Conversely, $m(q^*(v))$ is concave in v if $m'(q)$ is log-convex.
(1.) We focus on the negative-targeting case $A = [\underline{v}, v_1]$, but all arguments immediately extend to the case of positive targeting. Now consider the publisher's revenues as a function

of p.The total demand for advertising is given by

$$M(A)=\int_{\underline{v}}^{v_1} m(q^*(v))\,\mathrm{d}F(v)+(1-F(v_1))\,m(q^*(\bar{v}))\,.$$

Thus, we have

$$\begin{aligned}\frac{\partial M}{\partial v_1} &= (m(q^*(v_1))-m(q^*(\bar{v})))\,f(v_1)+(1-F(v_1))\,m'(q^*(\bar{v}))\frac{\partial q^*(\bar{v})}{\partial \bar{v}}\frac{\partial \bar{v}}{\partial v_1}\\ &= f(v_1)\left(m(q^*(v_1))-m(q^*(\bar{v}))+\frac{m'(q^*(\bar{v}))}{cm''(q^*(\bar{v}))}(\bar{v}-v_1)\right).\end{aligned}$$

This expression is positive if and only if $m''(q)/m'(q)$ is decreasing in q, *i.e.*, if $m(q^*(v))$ is convex. Because v_1 is decreasing in p, the publisher's revenue $c\cdot M$ is decreasing in p if $m'(q)$ is log-concave.

Now consider the data provider's revenues as a function of c. The inverse demand function $p(v_1)$ is given by

$$p(v_1)=\pi(v_1)-v_1 q^*([v_1,\bar{v}])+cm(q^*([v_1,\bar{v}]))\,.$$

For a fixed v_1, let $\hat{v}\triangleq\mathbb{E}[v\mid v>v_1]$, so that $q^*([v_1,\bar{v}])=q^*(\hat{v})$, with $v_1<\hat{v}$. Consider the derivative

$$\frac{\partial p(v_1,c)}{\partial c}=-(m(q^*(v_1))-m(q^*(\hat{v})))-(v_1-cm'(q^*(\hat{v})))\frac{\partial q^*(\hat{v})}{\partial c},$$

where

$$\frac{\partial q^*(\hat{v})}{\partial c}=-\frac{m'(q^*(\hat{v}))}{cm''(q^*(\hat{v}))}.$$

Using the first order condition $v=cm'(q^*(v))$ we obtain

$$\frac{\partial p(v_1,c)}{\partial c}=m(q^*(\hat{v}))-m(q^*(v_1))+(v_1-\hat{v})\frac{m'(q^*(\hat{v}))}{cm''(q^*(\hat{v}))}. \tag{25}$$

Notice that, as a function of v_1, the right-hand side of (25) is equal to zero if $v_1=\hat{v}$, and its derivative with respect to v_1 is equal to

$$\frac{\partial^2 p(v_1,c)}{\partial v_1\partial c}=-m'(q^*(v_1))\frac{\mathrm{d}q^*(v_1)}{\mathrm{d}v_1}+\frac{m'(q^*(\hat{v}))}{cm''(q^*(\hat{v}))}.$$

Because

$$\frac{\mathrm{d}q^*(v_1)}{\mathrm{d}v_1}=\frac{1}{cm''(q^*(v_1))}$$

we then obtain

$$\frac{\partial^2 p(v_1, c)}{\partial v_1 \partial c} = \frac{1}{c}\left(\frac{m'(q^*(\hat{v}))}{m''(q^*(\hat{v}))} - \frac{m'(q^*(v_1))}{m''(q^*(v_1))}\right). \tag{26}$$

Because $q^*(v)$ is strictly increasing in v, if $m''(q)/m'(q)$ is decreasing in q then the expression in (26) is positive, which implies $\partial p/\partial c$ is negative for all $v_1 < \hat{v}$. Therefore, if $m'(q)$ is log-concave, the inverse demand $p(v_1, c)$ is strictly decreasing in c, and so are the data provider's profits.

(2.) It is immediate to see that all results from part (1.) are reversed if $m'(q)$ is log-convex (so that $m''(q)/m'(q)$ is increasing in q and $m(q^*(v))$ is concave in v). ■

Appendix B: Information Structures

Information Structures Each advertiser i has a compact set $V_i = [0,1]$ of possible valuations for the contact with the customer, where a generic element is denoted by $v_i \in V_i$. The valuation v_i is independently distributed with prior distribution function $F(v_i)$, and the associated density function $g(v_i)$ is positive on V_i.

The signal space is denoted by $S_i \subseteq [0,1]$. The space S_i can either be countable, finite or infinite, or uncountable. Let $(V_i \times S_i, \mathcal{B}(V_i \times S_i))$ be a measurable space, where $\mathcal{B}(V_i \times S_i)$ is the class of Borel sets of $V \times S$. An *information structure* for advertiser i is given by a pair $\mathcal{S}_i \triangleq \langle S_i, F_i(v_i, s_i)\rangle$, where S_i is the space of signal realizations and $F_i(v_i, s_i)$ is a joint probability distribution over the space of valuations V_i and the space of signals S_i. We refer to this class of information structures as (Borel) measurable information structures. The joint probability distribution is defined in the usual way by

$$F_i(v_i, s_i) \triangleq \Pr(\widetilde{v}_i \leq v_i, \widetilde{s}_i \leq s_i).$$

The marginal distributions of $F_i(v_i, s_i)$ are denoted with minor abuse of notation by $F_i(v_i)$ and $F_i(s_i)$ respectively. For $F_i(v_i, s_i)$ to be part of an information structure requires the marginal distribution with respect to v_i to be equal to the prior distribution over v_i. The conditional distribution functions derived from the joint distribution function are defined in the usual way:

$$F_i(v_i \,| s_i) \triangleq \frac{\int_0^{v_i} dF_i(\cdot, s_i)}{\int_0^1 dF_i(\cdot, s_i)},$$

and similarly,

$$F_i(s_i \,| v_i) \triangleq \frac{\int_0^{s_i} dF_i(v_i, \cdot)}{\int_0^1 dF_i(v_i, \cdot)}.$$

The data provider can choose an arbitrary information structure $\mathcal{S}_i$ for every advertiser i subject only to the restriction that the marginal distribution equals the prior distribution of v_i. The cost of every information structure is identical and set equal to zero. The choice of $\mathcal{S}_i$ is common knowledge. At the interim stage every agent observes privately a signal s_i rather than her true match value v_i of the object. Given the signal s_i and the information structure $\mathcal{S}_i$ each advertiser forms an estimate about her true match value. The *expected value* of v_i conditional on observing s_i is defined as:

$$w_i(s_i) \triangleq \mathbb{E}[v_i \,| s_i] = \int_0^1 v_i dF_i(v_i \,| s_i).$$

Every information structure $\mathcal{S}_i$ generates a distribution function $H_i(w_i)$ over *posterior ex-*

pectations given by

$$H_i(w_i) = \int_{\{s_i : w_i(s_i) \leq w_i\}} dF_i(s_i).$$

We denote by W_i the support of the distribution function $H_i(\cdot)$. Observe that the prior distribution $F_i(\cdot)$ and the posterior distribution over expected values $H_i(\cdot)$ need not coincide. It is helpful to illustrate some specific information structures.

The information structure $\mathcal{S}_i$ yields perfect information if $F_i(v_i) = H_i(v_i)$ for all $v_i \in V_i$. In this case, the conditional distribution $F(s_i | v_i)$ has to satisfy

$$F_i(s_i | v_i) = \begin{cases} 0 & \text{if} \quad s_i < s(v_i), \\ 1 & \text{if} \quad s_i \geq s(v_i), \end{cases} \tag{27}$$

where $s(v_i)$ is an invertible function.

The information structure $\mathcal{S}_i$ is said to be *positively revealing* if

$$H_i(v) = \begin{cases} 0 & \text{if} \quad 0 < v \leq \widehat{w}_i; \\ F_i(\widehat{v}_i) & \text{if} \quad \widehat{w}_i \leq v \leq \widehat{v}_i; \\ F_i(v) & \text{if} \quad \widehat{v}_i \leq v \leq 1. \end{cases} \tag{28}$$

Thus, a positively revealing information structure implies $F_i(v_i) = H_i(v_i)$ for all $v_i \geq \widehat{v}_i \in V_i$, and pools all values $v < \hat{v}_i$ into the conditional expectation $\widehat{w}_i = \mathbb{E}[v_i | v_i \leq \widehat{v}_i]$.

An information structure $\mathcal{S}_i$ which satisfies (27) without necessarily satisfying the invertibility condition is called *partitional*. An information structure is called *discrete* if S_i is countable and *finite* if S_i is finite.

After the choice of the information structures $\mathcal{S}_i$ by the data provider, the induced distribution of the agent's (expected) valuations is given by $H_i(w_i)$ rather than $F_i(v_i)$. The signal s_i and the corresponding expected valuation $w_i(s_i)$ remain private signals for every agent i and the auctioneer still has to elicit information by respecting the truthtelling conditions.

Optimal Mechanism The data provider selects the information structures of the advertisers and a revelation mechanism. The objective of the data provider is to maximize his expected revenue subject to the interim participation and interim incentive constraints of the advertiser. The data provider can offer a menu of posterior expectations $H(w | \theta)$ at a price $t(\theta)$. In an incentive compatible mechanism the value function of an advertiser with a willingness to pay θ is given by

$$U(\theta) \triangleq \int_w \{\max\{\theta w - c, 0\} dH(w | \theta)\} - t(\theta), \tag{29}$$

and the interim incentive constraint requires that

$$\int_w \{\max\{\theta w - c, 0\} \, dH(w|\theta)\} - t(\theta) \geq \int_w \{\max\{\theta w - c, 0\} \, dH(w|\theta')\} - t(\theta'), \tag{30}$$

and the interim participation constraint requires that $U(\theta) \geq 0$.

For convenience, we shall restrict attention to a model with finite values and finite signals. We briefly discuss the extension to a continuum of types, values, and signal at the end. We present the finite model here as it allows to avoid additional qualification such as "almost surely" that arise in a model with a continuum of types, values or signals. A mechanism is then a transfer payment $t(\theta)$ and distribution $H_\theta : V \to \Delta(W)$ from values into expectations. We denote by $W(\theta)$ the set of posterior expectations under distribution H_θ:

$$W(\theta) \triangleq \{w \in W \,| h_\theta(w) > 0\}.$$

We say that $W(\theta)$ has binary support, in this case denoted by $B(\theta)$, if it contains only two elements:

$$B(\theta) = \{\underline{w}(\theta), \ \overline{w}(\theta)\} \tag{31}$$

and one of them leads to a contact, and the other one does not lead to contact: $\theta\underline{w}(\theta) - c < 0, \ \theta\overline{w}(\theta) - c \geq 0$.

Proposition 12 (Binary Mechanism)
Every optimal revenue mechanism can be implemented by a binary mechanism.

Proof. Consider an arbitrary, and finite, optimal mechanism $\{H_\theta, t(\theta)\}$. By hypothesis it satisfies the interim incentive constraints, that is for all types θ, θ', we have

$$U(\theta) \triangleq \left[\sum_{w \in W(\theta)} \max\{\theta w - c, 0\} \, h_\theta(w)\right] - t(\theta) \geq \left[\sum_{w \in W(\theta')} \max\{\theta w - c, 0\} \, h_{\theta'}(w)\right] - t(\theta') \triangleq U(\theta, \theta').$$

We denote by $W^+(\theta)$ the set of posterior expectations that lead to a contact with the advertiser, or

$$W^+(\theta) \triangleq \{w \in W \,| h_\theta(w) > 0 \wedge \theta w \geq c\},$$

and by $W^-(\theta)$ the set of posterior expectations that do not lead to a contact with the advertiser, or

$$W^-(\theta) \triangleq \{w \in W \,| h_\theta(w) > 0 \wedge \theta w < c\}.$$

Now, we can clearly bundle all the posterior expectations in $W^+(\theta)$ and in $W^-(\theta)$ to obtain a binary support as described in (31). Now clearly, under the constructed binary support, the indirect utility remains constants, but the value of a misreport is (weakly) smaller, that is for all $\theta \neq \theta'$:

$$\left[\sum_{w \in W(\theta')} \max\{\theta w - c, 0\} h_{\theta'}(w)\right] - t(\theta') \geq \left[\sum_{w \in B(\theta')} \max\{\theta w - c, 0\} h_{\theta'}(w)\right] - t(\theta'),$$

after all, in the original deviation the advertiser could have acted as in the binary support, but he had a possibly larger set of choices available to him, and hence is doing weakly worse in the binary mechanism, i.e. the value of a misreport has been (uniformly) lowered across all types. ■

By combining the posterior values into those with positive and those with negative value relative to the type θ of the agent, we do not change the value of the allocation for the agent. But, since the bundling/combination is performed with respect to the true type, it lowers the option value for all types other than the true type, because the binary mechanism forces them to take a constant action where before they might have chosen contingent actions. Thus, restricting the set of posterior realization only tightens the incentive constraints, and can only (weakly) improve the revenues for the principal.

Finally, Proposition 12 implies that every revenue-optimal mechanism can be implemented through a mechanism resembling cookie sales. In particular, we have the following corollary.

Corollary 1 (Noiseless Information Policy) *Every revenue-optimal mechanism can be implemented by a noiseless information policy.*

Intuitively, a noiseless information policy combines all posterior values with negative value into a single signal. As in Proposition 12, this does not change the value of the resulting allocation, but weakly lowers the value of any deviation. Therefore, the data provider can maximize revenues by revealing cookies above the optimal threshold in (14) and charging the corresponding prices.

References

ABRAHAM, I., S. ATHEY, M. BABAIOFF, AND M. GRUBB (2012): “Peaches, Lemons, and Cookies: Designing Auction Markets with Dispersed Information,” Discussion paper, Microsoft Research, Stanford, and MIT.

ADAMS, W., L. EINAV, AND J. LEVIN (2009): “Liquidity Constraints and Imperfect Information in Subprime Lending,” *American Economic Review*, 99(1), 49–84.

ADMATI, A., AND P. PFLEIDERER (1986): “A monopolistic market for information,” *Journal of Economic Theory*, 39(2), 400–438.

——— (1990): “Direct and indirect sale of information,” *Econometrica*, 58(4), 901–928.

ALLEN, F. (1990): “The market for information and the origin of financial intermediation,” *Journal of financial intermediation*, 1(1), 3–30.

ANDERSON, E. T., AND D. SIMESTER (2013): “Advertising in a Competitive Market: The Role of Product Standards, Customer Learning and Switching Costs,” *Journal of Marketing Research*, Ahead of Print.

ANTON, J., AND D. YAO (2002): “The sale of ideas: Strategic disclosure, property rights, and contracting,” *Review of Economic Studies*, 69(3), 513–531.

BABAIOFF, M., R. KLEINBERG, AND R. PAES LEME (2012): “Optimal mechanisms for selling information,” in *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 92–109. ACM.

BERGEMANN, D., AND A. BONATTI (2011): “Targeting in Advertising Markets: Implications for Offline Vs. Online Media,” *Rand Journal of Economics*, 42(3), 417–443.

BERGEMANN, D., AND M. PESENDORFER (2007): “Information Structures in Optimal Auctions,” *Journal of Economic Theory*, 137(1), 580–609.

BLAKE, T., C. NOSKO, AND S. TADELIS (2013): “Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment,” Discussion paper, eBay Research Labs, University of Chicago, and University of California Berkeley.

BUTTERS, G. (1977): “Equilibrium Distributions of Sales and Advertising Prices,” *The Review of Economic Studies*, 44(3), 465–491.

Celis, L., G. Lewis, M. Mobius, and H. Nazerzadeh (2012): "Buy-it-now or Take-a-chance: A simple sequential screening mechanism," Discussion paper, Harvard University.

García, D., and F. Sangiorgi (2011): "Information Sales and Strategic Trading," *Review of Financial Studies*, 24(9), 3069–3104.

Hoffmann, F., R. Inderst, and M. Ottaviani (2013): "Hypertargeting, Limited Attention, and Privacy: Implications for Marketing and Campaigning," Discussion paper, J.W. Goethe University Frankfurt, UCL, and Bocconi.

Hörner, J., and A. Skrzypacz (2012): "Selling Information," Discussion Paper CFDP 1743RR, Cowles Foundation, Yale University.

Iyer, G., and D. Soberman (2000): "Markets for product modification information," *Marketing Science*, 19(3), 203–225.

Kamenica, E., and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101(6), 2590–2615.

Kempe, D., V. Syrganis, and E. Tardos (2012): "Information Asymmetries in first-price common-value auctions, or, How to Sell Third-Party Cookies," Discussion paper, USC.

Mahdian, M., A. Ghosh, P. McAfee, and S. Vassilvitskii (2012): "To Match or Not to Match: Economics of Cookie Matching in Online Advertising," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 741–753. ACM.

Maskin, E., and J. Riley (1984): "Monopoly with Incomplete Information," *RAND Journal of Economics*, 15(2), 171–196.

Pancras, J., and K. Sudhir (2007): "Optimal marketing strategies for a customer data intermediary," *Journal of Marketing Research*, 44(4), 560–578.

Rayo, L., and I. Segal (2010): "Optimal Information Disclosure," *Journal of Political Economy*, 118(5), 949–987.

Rogerson, W. (2003): "Simple menus of contracts in cost-based procurement and regulation," *American Economic Review*, 93(3), 919–926.

Sarvary, M., and P. M. Parker (1997): "Marketing information: A competitive analysis," *Marketing Science*, 16(1), 24–38.

Sims, C. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3), 665–690.

Taylor, C. R. (2004): "Consumer privacy and the market for customer information," *RAND Journal of Economics*, 35(4), 631–650.

Xiang, Y., and M. Sarvary (2013): "Buying and selling information under competition," *Quantitative Marketing and Economics*, 11(3), 321–351.