SELLING IMPRESSIONS: EFFICIENCY VS. COMPETITION

By

Dirk Bergemann, Tibor Heumann, and Stephen Morris

July 2021

COWLES FOUNDATION DISCUSSION PAPER NO. 2291

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Selling Impressions: Efficiency vs. Competition*

Dirk Bergemann† Tibor Heumann‡ Stephen Morris§

July 27, 2021

Abstract

In digital advertising, a publisher selling impressions faces a trade-off in deciding how precisely to match advertisers with viewers. A more precise match generates efficiency gains that the publisher can hope to exploit. A coarser match will generate a thicker market and thus more competition. The publisher can control the precision of the match by controlling the amount of information that advertisers have about viewers. We characterize the optimal trade-off when impressions are sold by auction. The publisher pools premium matches for advertisers (when there will be less competition on average) but gives advertisers full information about lower quality matches.

JEL CLASSIFICATION: D44, D47, D83, D84.

KEYWORDS: Second Price Auction, Conflation, Digital Advertising, Impressions, Bayesian Persuasion, Information Design.

*We acknowledge financial support from NSF grants SES1824137 and SES 1948336, and ANID Fondecyt Iniciacion 11200365. We thank Steve Berry, Ozan Candogan, Phil Haile, Jason Hartline, Nicole Immorlica, Maxim Ivanov, Anton Kolotilin, Renato Paes Leme, Denis Shishkin, Rann Smorodinsky, Philipp Strack, Alex Wolitzky and Song Zuo for informative discussions. We thank seminar audiences at Informs AMD, UCSD, PUC Chile, the China Meetings of the Econometric Society and the Stony Brook International Conference on Game Theory for many helpful comments. While preparing the manuscript we become aware of Sorokin and Winter (2021) that establishes in Theorem 1 the result stated here in Proposition 1 with an independent and self-contained proof.

†Department of Economics, Yale University, New Haven, CT 06520, U.S.A., dirk.bergemann@yale.edu.

‡Instituto de Economía, Pontificia Universidad Católica de Chile, Santiago, Chile; tibor.heumann@uc.cl.

§Department of Economics, MIT, Cambridge, MA 02139, USA, semorris@mit.edu

1 Introduction

In the internet advertising market, it has become technologically feasible to match advertisers to viewers with ever-greater precision. But do publishers selling impressions have an incentive to do so? Finer matching generates efficiency gains that sellers can hope to exploit. But coarser matching generates market thickness, and so more market competition and less information rent for advertisers. Impressions are typically sold by auction and publishers can control the precision of the match by controlling the information that bidders have access to. We characterize the trade-off between efficiency and competition for the publisher.

We do this in two steps. First, we characterize what information a seller would choose to give buyers about their values in a second price auction in a standard independent private values setting. Second, we provide a model of the market for impressions in digital advertising markets, and show how our characterization applies in this setting. We now discuss these two steps in turn.

Consider a second price auction where bidders' valuations are independently and symmetrically distributed, but initially unknown to the bidders. The seller can choose what information each bidder can learn about their own value. If the seller did not allow them to learn anything, then all bidders would bid their (common) expected value and the good would be randomly (and inefficiently) allocated among them. If the seller allowed bidders to learn their true value, then they would have a dominant strategy (under private values in a second price auction) to bid their values. The good would be allocated efficiently to the bidder with the highest value. The revenue of the seller would equal the value of the efficient allocation minus the bidders' information rent. By permitting bidders to learn something but not everything about their values, the seller can trade off efficiency loss with information rent reduction. Our main result is a characterization of the optimal (among symmetric) information policies for the seller.

Conditional on having a low value, a bidder is likely to be competing with other bidders and earn low information rents. But conditional on having a high value, a bidder is likely to win (facing no competition) and thus can expect to win at a price significantly below his value, thus earning high information rents. Thus the gains from concealing information will be highest when valuations are high. In the optimal policy, high values are pooled and low values are revealed. There is a critical threshold described by a quantile above which all valuations are bundled together (Proposition 1). The threshold is given by a quantile of the distribution that depends only on the number of bidders (and not the distribution of valuations). The optimal quantile until which disclosure occurs is increasing in the number of participating bidders and goes towards 1 (i.e., full disclosure) as the number of bidders grows arbitrarily large. Thus, the information policy is influencing the distribution of bids, holding fixed the distribution of preferences among the bidders.

The assumption of a fixed finite number of bidders is extreme for many applications. We provide two results about what happens in large markets. First, we show that if the distribution of values has fat tails, then even as the number of bidders becomes large, and the quantile at which pooling starts approaches one, the gains from the optimal information policy relative to complete information or no information is high. We also consider the case where there is random entry into the auction. Suppose there is a prior probability that any bidder has a positive value for the object and a complementary probability that the bidder has zero value or no interest in the specific item. Then the optimal information structure intentionally invites the advertiser to bid on an item with positive probability even when the bidder has zero value for the item. That is, sometimes a market is made more competitive even when this lowers the expected value of the object (Proposition 3).

Our motivation for studying this problem is the market for impressions in digital advertising. A large share of digital advertising, whether in search, display advertising or social networks, is allocated by auction mechanisms. The second price auction is commonly used in digital advertising to form a match between competing advertisers (the *bidders*) and a viewer. A match between viewer and advertiser creates an *impression* (or search result) on the publisher's website. The seller (a *publisher* or *intermediary*) *or the publisher* by an auction platform sells the attention ("eyeball") of the viewer to competing advertisers. The viewer is thus the object of the auction. The viewers are typically heterogeneous in many attributes, their demographic characteristics, their preferences, their (past) shopping behavior, their browsing history and many other aspects, observable and unobservable. The advertisers therefore display a corresponding degree of heterogeneity in their willingness to pay for a match between their advertisement and a specific viewer. The private (and the social) value of any particular match is then determined jointly by a vector of attributes of the viewer and a vector of preferences for those attributes of the advertiser. In the presence of this heterogeneity on both sides of the match, viewer and advertiser, internet advertising has moved towards targeted advertising to join the information. The auction can therefore support highly targeted advertising that may increase the social efficiency in the match formation between viewer and advertiser. But - as discussed earlier - allowing for finely targeted bidding may also thin the market among the advertisers, and hence reduce the competition between advertisers. Publishers distinguish two schemes, or algorithms, for mapping preferences and attributes into bids, *automated bidding* and *manual bidding.* In automated bidding, *autobidding* for short, the seller offers a bidding algorithm that generates optimal bids for the advertisers given the disclosed information.[1] In manual bidding, the seller offers a disclosure

[1]See Google Ads Help Center (2021b), Google Ads Help Center (2021a) or Facebook Business Help Center (2021) for summary descriptions of automated bidding mechanisms.

algorithm that generates information about the attributes, and in particular a bid recommendation, which each bidder then manually adopts or modifies into a bid for the impression, also referred to as dashboard mechanism in Hartline, Johnsen, Nekipelov, and Zoeter (2019). Autobidding has become increasingly prevalent in digital advertising to convert the high-dimensional information across millions of impressions into bids with minimal latency, see Aggarwal, Badanidiyuru, and Mehta (2019) and Deng, Mao, Mirrokni, and Zuo (2021).

In this market, publishers can control the information that advertisers have about their values. While there are many reasons why advertisers' values might be correlated, they will not be if variation in viewers' attributes is horizontal. A second contribution of the paper is to develop a stylized model of the market for impressions, establish that our earlier model applies to this market, and show how the results apply under reasonable assumptions on the market.

In our model of this market, a viewer is characterized a (perhaps high-dimensional) attribute. The publisher (but not the advertisers) knows the attribute of the viewer. The advertiser (but not the publisher) knows his preference over attributes. The viewer attribute can be combined with an advertiser's preference to generate a match quality, and the value of the viewer to the advertiser is an increasing function of the match quality. A key feature of this model is that the advertiser's private information is not informative about his value of a viewer unless it is combined with information held by the publisher. The publisher sells impressions (the allocation of the viewer to an advertiser) in a second price auction. Because only the publisher knows the attributes of the viewer, the publisher can control the information that advertisers have about their match quality with the viewer.

We present two results in our model. We first show that this model of two-sided information gives rise to the setting of our main result: an independent private values setting, where the advertiser has no information about his value of the viewer, but the combination of the publisher's information about attributes and advertiser's information about preferences fully reveals the state. Now we can identify the optimal information structure from our earlier analysis. But how can this be implemented in practise? We consider *autobidding*. The advertiser reports his preferences to the publisher and the publisher commits to bid optimally for the bidder as a function of the optimal information structure. We show that the bidder will have an incentive to truthfully report his preferences. Thus autobidding implements the optimal outcome in this market.

Literature Levin and Milgrom (2010) suggested that the idea of conflation (central in many commodity markets) by which similar but distinct products are treated as identical in order to make markets thicker or reduce cherry-picking, may be relevant for the design on online advertising markets. The information structure (4) determines exactly when conflation should occur, in the upper interval, and when not, in the lower interval.

The paper relates to the literature studying optimal information disclosure in selling mechanisms. Ganuza (2004) studies the optimal information disclosure in a second-price auction, where bidder's valuation are determined by the quality of a match between a bidder's taste and the good's characteristic represented by Hotelling model on a circle. The seller chooses the optimal public signal about the good's characteristic. He shows that the equilibrium information provision is less than the surplus-maximizing one. His result is about *costly* public signals and so it does not address the trade-off between efficiency and market thickness that is central in our paper.

Bergemann and Pesendorfer (2007) analyze the joint optimal design of auction and information structure. In particular, they allow for asymmetric information structures and personalized reserve prices. Here, we fix the selling mechanism to be a second-price auction as it is a better fit for the markets for selling impressions.

Palfrey (1983) studies the bundling decision of a monopolist who sells J goods to N bidders via a second-price auction. He gives necessary and sufficient conditions for a single bundle to generate a higher revenue than J independent auctions. The trade-off that governs the decision is similar as in our model, efficiency vs. market thickness, but the decision available for the seller is coarser as he compares only two options, to bundle or not to bundle. This is akin to comparing when no information generates higher revenue than complete information in our model.

Similar high-dimensional models with attribute (or features) and preferences have appeared recently in dynamic pricing literature. Here, the seller does not initially know the values of the different features, but can learn the values of the features based on whether products were sold at the posted prices in the past, see Cohen, Lobel, and Leme (2020).

2 Model

There are N agents who bid for an indivisible good in an auction. Bidder i's valuation is denoted by v_i. We assume that the valuations are independently and identically distributed across agents according to an absolutely continuous distribution, denoted by F. The assumptions that F is absolutely continuous helps simplify some of the expressions but all results go through unchanged if we relax this assumption.

The seller can choose how much information each bidder will have about his own valuation. An information structure is denoted by:

$$s_i : \mathbb{R}_+ \to \Delta\mathbb{R}_+,$$

where $s_i(v_i)$ is the signal observed by bidder i when his valuation is v_i. After observing s_i, the bidder forms his beliefs about his valuation. An agent's expected valuation is denoted by:

$$w_i \triangleq \mathbb{E}[v_i \mid s_i].$$

We denote by G_i the distribution of expected valuations. Note that we are making two assumptions about the information structure. First, each bidder only observes information about his own valuation, which is reflected by the fact that s_i takes as an argument v_i only (instead of $(v_1, ..., v_N)$). Additionally, there is no common source of randomization in the signals. Hence, the signals will be independently distributed across agents. Finally, we assume that the seller is restricted to symmetric information structures, i.e., $s_i(\cdot) = s_j(\cdot)$.

The objective of the seller is to maximize revenue. Since agents are bidding in a second-price auction it is a dominant strategy to bid their expected valuation. Hence, revenue is equal to the second-highest expected valuation across bidders. We denote the k-th highest valuation by $w_{(k)}$. The objective of the seller is to solve:

$$R \triangleq \max_{\{s:\mathbb{R}\to\Delta\mathbb{R}\}} \mathbb{E}[w_{(2)}]. \tag{1}$$

3 Optimal Information Structure

Since the expected revenue is equal to the expectation of second-highest valuation, the distribution of expected valuations generated by the signal is a sufficient statistic to compute the seller's expected revenue. Hence, instead of studying explicitly the signal chosen by the seller, we frequently refer to the distribution of expected valuations generated by a signal (recall that this is denoted by G).

The second-order statistic of N symmetrically and independently distributed random variables is distributed according to

$$\mathbb{P}(w_{(2)} \leq t) = NG^{N-1}(t)(1 - G(t)) + G^N(t).$$

The expected revenue of the auctioneer is therefore:

$$\mathbb{E}[w_{(2)}] = \int_0^\infty td(NG^{N-1}(t)(1 - G(t)) + G^N(t)).$$

We now characterize the set of feasible distributions G.

By Blackwell (1951), Theorem 5, there exists a signal s that induces a distribution of expected valuations if and only if F is a mean preserving spread of G. F is defined to be a mean preserving spread of G if

$$\int_v^\infty dF(t) \leq \int_v^\infty dG(t),\ \forall v \in \mathbb{R}_+$$

and

$$\int_0^\infty dF(t) = \int_0^\infty dG(t).$$

If F is a mean preserving spread of G we write $F \prec G$.

We can now express the seller's problem as finding an optimization over a distribution G subject to a mean-preserving restriction. The choice of the optimal information structure can be written as the following maximization problem:

$$R = \max_G \int_0^\infty td(NG^{N-1}(t)(1-G(t)) + G^N(t)) \tag{2}$$
$$\text{subject to } F \prec G.$$

This problem consists of maximizing over feasible distributions of expected valuations. However, the objective function is non-linear in the probability (or density) of the optimization variable G. Moreover, the non-linearity cannot be confined to be either concave or convex on G.

The key step in our argument comes from a change of variables, re-writing the above in terms of the quantile q of the second order statistic. We denote by $S_N(q)$ the cumulative distribution function of the quantile of the second-highest valuation: $S_N(q) \triangleq \mathbb{P}(F(w_{(2)}) \leq q)$. We index by N to highlight the dependence on the number of buyers. We observe that $S_N(q)$ is given by:

$$S_N(q) = Nq^{N-1}(1-q) + q^N.$$

The quantile distribution S_N is independent of the underlying distribution F or G. Just as the quantile of any random variable is uniformly distributed, the quantile of the second-order statistic of N symmetric independent random variables is distributed according to S_N for any underlying distribution. Hence, the revenue can be computed by taking the expectation over quantiles using measure $S_N(q)$: the revenue given the quantile of the second-order statistic is G^{-1}. So maximization problem (2) can be transformed into:

$$\max_{G^{-1}} \int_0^1 S_N'(q) G^{-1}(q) dq \tag{3}$$
$$\text{subject to } G^{-1} \prec F^{-1}$$

The corresponding constraint states that the seller can choose any distribution of expected valuations whose quantile function G^{-1} is a mean-preserving spread of the quantile function F^{-1} of the initial distribution of valuations. This uses a well-known property of the distribution function, see Shaked and Shanthikumar (2007), Chapter 3, stating that $F \prec G$ if and only if $G^{-1} \prec F^{-1}$. Hence, we have a linear (in G^{-1}) maximization problem subject to a majorization constraint, which will allow us to solve the problem with known methods.

Proposition 1 (Optimal Information Structure)
The unique optimal symmetric information structure is given by:

$$s(v_i) = \begin{cases} v_j & \textit{if } F(v_i) \leq q_N^* \\ \mathbb{E}[v_j \mid F(v_j) \geq q] & \textit{if } F(v_i) \geq q_N^* \end{cases} \tag{4}$$

where the critical quantile $q_N^* \in [0,1)$ is independent of F. In particular, $q_2^* = 0$; q_N^* is increasing in N; $q_N^* \to 1$ as $N \to \infty$; and for each $N \geq 3$, q_N^* is the unique solution in $(0,1)$ to:

$$S_N'(q)(1-q) = 1 - S_N(q). \tag{5}$$

Note that (5) is an $N-$th degree polynomial in q. Thus, the optimal information structure is to reveal the valuation of all those bidders who have a valuation lower than some threshold determined by a fixed quantile q_N^* and otherwise reveal no information beyond the fact that the valuation is above the threshold. The threshold in terms of the valuation is given by $F^{-1}(q_N^*)$, but the quantile q_N^* is independent of the distribution F of valuations. The optimal information structures thus supports more competition at the top of the distribution at the expense of an efficient allocation. The information structure bundles for every bidder all valuations above the threshold $F^{-1}(q_N^*)$ into a single mass point. It therefore fails to distinguish in the allocation between any two valuations that are in the upper tail of the distribution $[F^{-1}(q_N^*), \infty)$. The benefit accrues through more competitive bids among the high value bidders. Namely, if the second highest bid is in the interval, then its competitive bid matches exactly the bid of the winning bid, and thus the information rent of the winning bidder is depressed considerably with a corresponding gain in the revenue for the seller.

The pooling at the top ensures competition when values are high. A natural question to ask is how many bidders will be in the pooling region. Note that the number of bidders with valuations above the threshold has a binomial distribution with parameters $(N, 1-q_N^*)$. We can evaluate numerically the expected number of bidders who have values above the threshold: It is always in the interval $[1.75, 2.25]$

and is decreasing in N for $N \geq 3$. Results in the following Section imply that the expectation converges down to 1.79 as $N \to \infty$.

Before we prove the result, we will provide some intuition for the critical quantile by confirming that it must be given by equation (5) as long as information takes the form given in the proposition. Suppose that we fix a quantile threshold q and write $v = F^{-1}(q)$ for the corresponding value. At the cutoff v there is a discontinuity in the bids as a function of the bidders' valuation. Bidders with values marginally below v will bid essentially v; bidders with values marginally above v will bid $\mathbb{E}_F[t|t \geq v]$. The difference is given by:

$$\Delta \triangleq \mathbb{E}_F[t|t \geq v] - v.$$

Now what happens to revenue if we decrease the threshold by dq? Now with probability $S_N'(q)\, dq$ the second-highest bid was not in the pooling zone before the decrease and is after the decrease; and revenue increases by Δ. With probability $1 - S_N(q)$, the second-highest bid was in the pooling zone before the decrease and there is a loss of revenue of $\frac{d}{dq}\Delta$. But one can show that $\frac{d}{dq}\Delta = \frac{1}{1-q}\Delta$, so equating expected loss and gains requires

$$S_N'(q)\, dq\Delta = (1 - S_N(q)) \frac{1}{1-q} \Delta dq$$

giving (5).

To prove Proposition 1, we state a result of Kleiner, Moldovanu, and Strack (2021) in terms of our maximization problem (3).

Proposition 2 (Kleiner, Moldovanu, and Strack (2021), Proposition 2) *Let G^{-1} be such that for some countable collection of intervals $\{[\underline{x}_i, \bar{x}_i) \mid i \in I\}$,*

$$G^{-1}(q) = \begin{cases} F^{-1}(q) & q \notin \cup_{i \in I}[\underline{x}_i, \bar{x}_i) \\ \frac{\int_{\underline{x}}^{\bar{x}} F^{-1}(t)dt}{\bar{x}_i - \underline{x}_i} & q \in [\underline{x}_i, \bar{x}_i) \end{cases}$$

If $\text{conv}S_N$ *is affine on* $[x_i, x_i)$ *for each* $i \in I$ *and if* $\text{conv}S_N = S_N$ *otherwise, then* G^{-1} *solves problem* (3). *Moreover, if* F^{-1} *is strictly increasing the converse holds.*

Here, $\text{conv}S_N$ is the convexification of S_N, i.e., the largest convex function that is smaller than S_N. With this result we can prove our main result.

Proof of Proposition 1. The second derivative of the distribution S_N of the quantile of the second order statistic is given by:

$$S_N''(q) = q^{N-3}(N-1)N(N-2-q(N-1)).$$

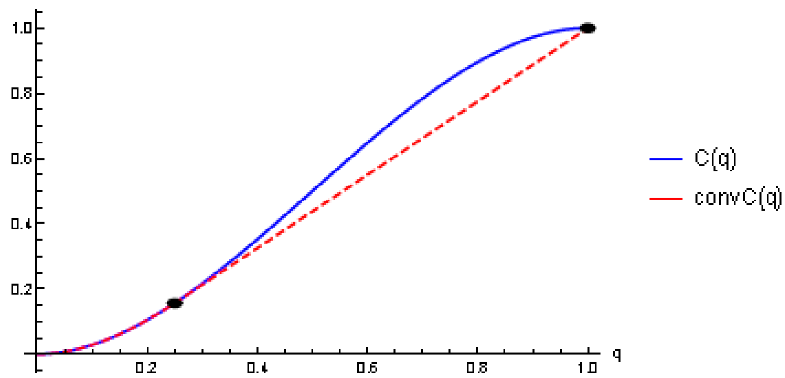

Figure 1: Convexification of $S(q)$ for $N = 3$.

Hence, $S_N(q)$ is concave if and only if

$$q \geq (N-2)/(N-1),$$

and convex otherwise. Thus, the convex hull of S_N for $N \geq 3$ is given by:[2]

$$\text{conv}S_N\,(q) = \begin{cases} S_N(q) & \text{if } q \leq q_N^*; \\ S_N'(q_N^*)(q-q_N^*) + S(q_N^*) & \text{otherwise.} \end{cases}$$

where q_N^* is defined as in (5) for $N \geq 3$. In Figure 1 we illustrate S_N and convS_N for $N = 3$.

For $N = 2$, we have

$$\text{conv}S_2\,(q) = q$$

and can define $q_2^* = 0$. Now let G^{-1} be given by:

$$G^{-1}(q) = \begin{cases} F^{-1}(q) & q < q_N^* \\ \frac{\int_{q_N^*}^1 F^{-1}(t)dt}{1-q_N^*} & q \in [q_N^*, 1) \\ 1 & q = 1 \end{cases} \tag{6}$$

Then, G^{-1} satisfies all the assumptions of Proposition 2, so it is the unique optimal solution to (3). For all valuations below $G^{-1}(q_N^*)$ the distribution over expected valuations is the same as that of the real

[2] To verify this is the convex hull, note that $\tilde{q} \leq (N-2)/(N-1)$ so by construction conv$S$ is convex for $\tilde{q} \leq (N-2)/(N-1)$, affine for $\tilde{q} \geq (N-2)/(N-1)$, and with continuous derivative, so it is convex. Also, by construction, whenever conv$S < S$ (i.e., the affine section), the graph of conv$S$ is in the graph of the convex hull of S.

valuations. Hence, types below $G^{-1}(q_N^*)$ know their own values. On the other hand, for valuations above $G^{-1}(q_N^*)$ the distribution over expected valuations is a mass point at the expected valuation conditional on being above $G^{-1}(q_N^*)$. Hence it is clear that this distribution is induced by information structure (4).

To check that q_N^* is strictly increasing in N we define:

$$\psi(q, N) \triangleq S_N'(q)(1-q) - (1 - S_N(q)).$$

By definition, $\psi(q_N^*, N) = 0$. We now note that:

$$\psi(q, N+1) - \psi(q, N) = N(q-1)^2(N(q-1)+1)q^{N-2}.$$

so $\psi(q, N+1) - \psi(q, N) \geq 0$ if and only if $q \geq (N-1)/N$. As previously argued, $q_N^* < (N-2)/(N-1)$ so $q_N^* < (N-1)/N$, which implies that:

$$\psi(q_N^*, N+1) < 0. \tag{7}$$

We also have that $\psi(0, N) = -1$ and $\psi(1-\varepsilon, N) > 0$ for ε small enough, where the last part can be verified by noting that

$$\psi(1, N) = \frac{\partial \psi(1, N)}{\partial q} = 0 \quad \text{and} \quad \frac{\partial^2 \psi(1, N)}{\partial q^2} = N(N-1) > 0.$$

As previously argued $\psi(q, N+1)$ has a unique root in $(0, 1)$, so (7) implies that $q_N^* < q_{N+1}^*$.

Finally, if N diverges to infinity and $\lim_{N\to\infty} q_N^* < 1$, then in the limit we would have that

$$S_N(q_N^*), S_N'(q_N^*) \to 0.$$

So (5) would not be satisfied. We thus must have that $\lim_{N\to\infty} q_N^* = 1$. ■

The information structure (4) that emerges here for every bidder is sometimes referred to as "upper censorship" in the Bayesian persuasion literature, as it pools all the states above a cutoff and reveals all the states below the cutoff, see Proposition 3 in Alonso and Camara (2016) or Theorem 1 in Kolotilin, Mylovanov, and Zapechelnyuk (2021). It is useful to compare our problem to a Bayesian persuasion model where the objective function permits a nonlinear evaluation $u(x)$ of an outcome but linear in the probability, thus

$$\max_G \int_0^1 u(t) dG(t),$$
$$\text{subject to } F \prec G,$$

as, for example in Dworczak and Martini (2019). Our original maximization problem (2) did not take this form as it was non-linear in probabilities. However, we reformulated the problem to one that is linear in the new optimization variable G^{-1}, changing the direction of the constraint. For this problem, the convexification of S_N was key to identifying the optimal information structure. The fact that S_N is always convex-concave then generated the upper censorship information structure.

4 Large Markets

We now develop some implications of the optimal information structure in markets with a large number of (possible) bidders which is arguably the prevailing condition in digital advertising. We first consider how the information responds to the random participation of bidders. We then consider the revenue performance of the auction with the optimal information structure when the actual number of participating bidders grows large. Here we analyze the a class of distribution with heavy tails that have been proposed by Arnosti, Beck, and Milgrom (2016) in their analysis of internet advertising.

Random Number of Bidders We now assume that the valuation of bidders is distributed according to F with probability $1-p$ and with probability p the valuation is equal zero. To analyze the optimal information structure, we analyze the optimal quantile in the limit as $N \to \infty$ and $p \to 1$. We keep the expected number of bidders who have strictly positive values constant at:

$$\lambda \triangleq N(1-p).$$

In this limit, the number of bidders who have a strictly positive values is randomly distributed according to a Poisson distribution with parameter λ (this is called the law of rare events or the Poisson limit theorem). Hence, in the limit, it is as if the number of bidders is randomly distributed.

To characterize the optimal information structure, let ρ be the unique solution larger than zero to the following equation:

$$\rho^2 + \rho + 1 = e^{\rho}, \tag{8}$$

with $\rho \approx 1.793$. We denote the expected valuation conditional on v being drawn from the distribution F by

$$v_F = \mathbb{E}_F\left[v\right].$$

Proposition 3 (Optimal Information)
In the limit as $N \to \infty$, $p \to 1$, the optimal information structure is:

1. *If $\lambda \leq \rho$, then bidders observe binary signals and their expected valuation is either* 0 *or* $v_F \lambda/\rho$.

2. *If $\lambda > \rho$, bidders with valuation $F(v_i) \leq (\lambda - \rho)/\lambda$ learn their value, and bidders with valuation $v_i \in [F^{-1}((\lambda - \rho)/\lambda), 1]$ only learn that their valuation is in this interval.*

Proof of Proposition 3. For any fixed N, we define the expected number of bidder who have value above the q_N^* quantile as:

$$\rho \triangleq N(1 - q_N^*). \tag{9}$$

In the limit $N \to \infty$, (5) converges to the following equation (in terms of ρ):

$$\rho^2 e^{-\rho} = 1 - e^{-\rho} - \rho e^{-\rho},$$

We then get the result by applying Proposition 1. ■

The distribution of bidders whose expected valuation is above the cutoff quantile is distributed according to a Poisson distribution with parameter ρ (regardless of λ):

$$\rho = N(1 - q_N^*), \tag{10}$$

So the expected number of bidders whose expected valuation is above the cutoff quantile is equal to ρ. It is also interesting that the probability that there is just one bidder above the cutoff is approximately 0.3, while the probability that there are at least 3 bidders is approximately 0.27. In the former case, there is not enough competition to extract this bidder's surplus, in the latter case there are excess bidders so it would be best to have less bidders with higher valuation. The optimal information structure approximately equates the probability of these two type of errors.

In the model with a random number of bidders, we also get a more nuanced analysis when there are few bidders. We can see that if expected number of "serious" bidders (i.e., bidders with non-zero valuation) is small, it is optimal to attract bidders whose valuation is zero and disclose no information. This is intuitive, when there are few bidders the priority is increasing market thickness, which comes about at the cost of lower expected valuations. In terms of the random entry, this means that it is part of the optimal information structure to present objects to the bidder that have zero value to the bidder as long as the bundle also includes objects that have positive value, but the bidder is bidding for them without being able to make the distinction. This practice thus bundles low value and high value impression to maintain a competitive market.

The optimal information structure thus supports a match process between advertiser and viewer that is often referred to as "broad match" in digital advertising, see Dar, Mirrokni, Muthukrishnan, Mansour, and Nadav (2009) and Eliaz and Spiegler (2016). By broad matching which is common in ad auctions, the seller provides matches not only on the exact and narrow matches for specific keywords and characteristics but also for a larger, hence broad set of matches, see Google Ads Help Center (2021c). Proposition 3 then establishes that broad matching is an important instrument to maintain competition in the ad auction. The first part of the Proposition 3 suggests that with thin markets, it is even optimal to include irrelevant matches with value 0 and thus lower the expected value as it increases the probability of competitive bids.

Revenue Performance with Large Number of Bidders We now examine the difference between the revenue generated under the optimal information structure and under complete information when the number of bidders become large. Throughout this subsection we assume that the density of distribution of valuations has regularly varying tails with index $\alpha - 1$, that is,

$$\lim_{t \to \infty} \frac{f(kt)}{f(t)} = k^{\alpha-1}, \text{ for all } t > 0.$$

We assume that $\alpha < -1$.[3] For example, the Pareto distribution satisfies this assumption, $f(v) = -\alpha v^{\alpha-1}$ for $v \geq 1$. Intuitively, the densities that have regularly varying tails have fat tails that decay as a Pareto distribution with shape parameter α. As argued by Arnosti, Beck, and Milgrom (2016), the Pareto distribution, with its fat tail provides a good fit for demand in the digital advertising market, our leading application.

We denote by R_c the expected revenue in the second price auction under complete disclosure of information:

$$R_c \triangleq \mathbb{E}[v_{(2)}].$$

We now compare the revenue under the optimal information structure, R with the revenue under complete disclosure, R_c for large N.

Proposition 4 (Revenue Gain with Large Number of Bidders)
As the number of bidders grows, there exists $z \in (1, \infty)$ *such that:*

$$\lim_{N \to \infty} \frac{R}{R_c} = z. \tag{11}$$

Furthermore, in the limit $\alpha \to -1$, $z \to \infty$.

As the number of bidders grows, the gains from using an optimal information structure does not vanish. The reason is that despite there being many bidders, there always remains a sufficiently high probability

[3]The condition $\alpha < -1$ is necessary and sufficient to guarantee that the distribution of valuations has finite first moment.

of there being a bidder with a disproportionately high valuation. Under complete information, it is not possible to extract the surplus from this bidder because there is not enough competition. Formally, when the distribution has fat tails:[4]

$$\mathbb{E}[v_{(1)}] - \mathbb{E}[v_{(2)}] \to \infty, \text{ as } N \to \infty. \tag{12}$$

By contrast, the optimal information structure thickens the market at the tail of the distribution and can thus provide a revenue improvement even as the numbers of bidders becomes arbitrarily large.

5 Market for Impressions

We now translate the earlier results into a market for impressions with two-sided information. This translation recasts the optimal information design as bidding mechanisms in the world of digital advertising. The choice of the optimal information structure can then interpreted in terms of the information policy of the publisher who matches the viewer with the advertisers.

The viewer has attribute $x \in X \subset \mathbb{R}^J$ distributed according to F_x. Each advertiser i has a preference for the attributes described by, $y \in Y \subset \mathbb{R}^J$, distributed according to F_y, identically and independently distributed across advertisers.

An impression is a match between an advertiser and a viewer. The value v_i of advertiser i from attracting a viewer is determined by a function u:

$$u : X \times Y \to \mathbb{R}_+,$$

such that:

$$v_i \triangleq u(x, y),$$

and we refer to u as the *valuation function*. The distribution of characteristics (x, y) and the valuation function u induce a distribution of the bidder i's value v_i, which we denote by F.

We assume that the unconditional distribution of values $(v_1, ..., v_N)$ generated by (x, y) are independent across bidders. We further assume that the unconditional distribution of values are the same as the conditional distribution of values $(v_1, ..., v_N)$ conditional on either x or y. That is, $(x, v_1, ..., v_N)$ and $(y, v_1, ..., v_N)$ are random vectors consisting of independently distributed random variables. Of course, (x, y_i, v_i) are *not* independently distributed. In other words, the preference vector y_i provides only information about the value v_i when combined with information about the attributes x of the viewer. Moreover, each advertiser

[4] This result is immediately implied by the analysis in the proof of Proposition 4.

i's belief about the valuation of their competitors is unaffected by attribute x or preference y_i. Thus, the present model of attributes and preferences generates a model of independent private values.

These assumptions implicitly impose restrictions on the set of valuations functions and distributions of attributes and characteristics that we consider. In other words, not every valuation function u and distributions F_x, F_y will generate joint distributions of values that satisfy these assumptions. The attributes of viewer therefore reflect an aspect of horizontal differentiation with value implications that depend on the preferences of the advertiser. Conversely, an aspect of vertical differentiation that has similar implications across all preferences would fail the equivalence between unconditional and conditional value distribution.

We now briefly describe two classes of models that satisfy the above conditions, one high-dimensional and one low-dimensional set of models.

Our leading example is given by the following specification. Let there be a vector of attributes $x \in \{-1,1\}^J$, so attribute j takes values -1 or 1. Each advertiser i has a preference for attribute j, $y_{ij} \in \{-1,1\}$. Thus $y_i \in \{-1,1\}^J$ is advertiser i's preference, and $y = (y_1, ..., y_N) \in \{-1,1\}^{NJ}$ is the profile of the preferences of the advertisers. The attributes and preferences are uniformly and independently distributed across components and bidders. The valuation function is given by:

$$u(x, y_i) = u\left(\frac{1}{\sqrt{J}}\sum_{j=1}^{J} x_j y_{ij}\right), \tag{13}$$

for some strictly increasing function u.

An alternative class of models is given the following Hotelling location model that has only a one-dimensional space of uncertainty. Suppose that $x, y_i \in [0,1]$ are positions in a circle of perimeter 1 uniformly and independently distributed. Let d_i be the shortest distance between x, y_i on the circle and

$$u(x, y_i) = u(d_i(x_i, y)) \tag{14}$$

for some strictly decreasing function u.

Both of these classes of models satisfy the independence conditions above.

We will now analyze automated bidding in the second-price auction based on a signal of the value of the impression. Thus, the publisher commits: (i) to complement the advertiser's information with a signal regarding the match quality; and (ii) to set the advertiser-optimal bid. In turn, the advertiser submits his preference y (and thus a description of the attributes he cares about). The central aspect of automated bidding is that the publisher complements the advertiser's private information y with information about the viewer's attribute x that is unknown to the advertiser.

Formally, the publisher chooses a signal

$$s_i : \mathbb{R} \rightarrow \Delta\mathbb{R}$$

as a function of the advertiser's reported value $v_i(x, y_i)$. While the publisher cannot directly observe v_i, she elicits the advertiser's preferences and knows the viewer attributes, so she can infer v_i. The publisher submits a bid $b_i : Y \times \mathbb{R} \rightarrow \mathbb{R}$ satisfying that:

$$b_i(y_i, s_i) = \mathbb{E}[v_i \mid y_i, s_i(v_i)] \tag{15}$$

That is, the publisher submits a bid on behalf of advertiser i equal to the advertiser's expected valuation given his preferences y_i and the additional information s_i provided by the publisher. This is a advertiser's dominant strategy given the available information $(y_i, s(v_i))$. Because (v_i, y_i) are independently distributed, we have that:

$$\mathbb{E}[v_i \mid y_i, s_i(v_i)] = \mathbb{E}[v_i \mid s_i(v_i)].$$

Hence, the publisher's problem reduces to find an information structure s that solves the original problem stated earlier in (1):

$$R \triangleq \max_{s:\mathbb{R}\rightarrow\Delta\mathbb{R}} \mathbb{E}[b_{(2)}].$$

We now verify that under the automated bidding in the second price auction, it is optimal for each advertiser to truthfully report their preferences to the publisher. A reporting strategy for bidder i is denoted by:

$$\widetilde{y}_i : Y \rightarrow \Delta Y.$$

Given the reported preferences, the seller discloses to the bidder a signal $s(\widetilde{v}_i)$, where $\widetilde{v}_i$ is the valuation of an advertiser with preferences $\tilde{y}_i$. The induced bid is denoted by $\tilde{b}_i$, while b_i denotes the bid when preferences are reported truthfully.

Proposition 5 (Truthful Reporting)
Under the optimal information structure, it is a dominant strategy for an advertiser to report truthfully his preferences to the publisher.

Proof of Proposition 5. By assumption, (v_i, y_i) are independently distributed. Thus the distribution of the bid $\tilde{b}_i$ is the same for every reported preference $\widetilde{y}_i$. Of course, the joint distribution of $(v_i, \tilde{b}_i)$ does depend on the reporting strategy. In fact, we note that for all $v', b' \in \mathbb{R}$:

$$\Pr(v_i \leq v', \tilde{b}_i \leq b') \leq \min\{\Pr(v_i \leq v'), \Pr(\tilde{b}_i \leq b')\} = \min\{\Pr(v_i \leq v'), \Pr(b_i \leq b')\} = \Pr(v_i \leq v', b_i \leq b'),$$

The first inequality is true for any bivariate distribution, the first equality is because both distributions have the same marginals, and the second equality follows from the fact that:

$$\Pr(v_i \le v', b_i \le b') = \begin{cases} F(v') & \text{if } v' \le b' \text{ or } b' = \mathbb{E}[v_i \mid v_i \ge F^{-1}(q_N^*)], \\ F(b') & \text{otherwise.} \end{cases}$$

By definition, this means that (v_i, b_i) is greater than $(v_i, \tilde{b}_i)$ in the positive quadrant dependent order (see Shaked and Shanthikumar (2007)).

We now write the difference in the expected bidder's surplus under truthful reporting and misreporting as follows:

$$\begin{aligned} \Delta V &\triangleq \mathbb{E}[(v_i - b_i)\Pr(b_i \ge \max_{j \ne i} b_j)] - \mathbb{E}[(v_i - \tilde{b}_i)\Pr(\tilde{b}_i \ge \max_{j \ne i} b_j)] \\ &= \mathbb{E}[(v_i \Pr(b_i \ge \max_{j \ne i} b_j)] - \mathbb{E}[v_i \Pr(\tilde{b}_i \ge \max_{j \ne i} b_j)], \end{aligned}$$

where $\Pr(b \ge \max_{j \ne i} b_j)$ is the probability that the $N-1$ competing bids are less than b and we use that the distribution of b_i is the same as $\tilde{b}_i$ to cancel out two terms. Clearly $\Pr(b \ge \max_{j \ne i} b_j)$ is an increasing function of b, so $v \Pr(b \ge \max_{j \ne i} b_j)$ is supermodular in (v, b), which implies that

$$\mathbb{E}[v_i \Pr(\tilde{b}_i \ge \max_{j \ne i} b_j)] \le \mathbb{E}[v_i \Pr(b_i \ge \max_{j \ne i} b_j)],$$

see (9.A.18.) in Shaked and Shanthikumar (2007). ■

The proposition states that advertisers are willing to submit their preferences honestly to the publisher. The intuition for the proof is that misreporting would not change the distribution of bids, but instead, only decrease the correlation between an advertiser's bid and his valuation. The automated bidding algorithm can alternatively be interpreted as a restriction on the bidding language imposed by a publisher. The publisher then runs a second price auction conditional on the realized attributes and the bidding rules.

We can illustrate this in our leading example of binary characteristics described above in (13). The optimal information structure can be implemented in a straightforward manner in the model of characteristics. Namely, the seller informs each bidder about the number of matched characteristics as long as this number is smaller than a threshold number n^* implied by the optimal quantile q_N^*. If the number of matched characteristics exceeds n^*, then the seller only reports that the realized matches exceed the threshold number n^*. Thus, a restriction in terms of the bidding language would allow the bidder to place bids as a function of the matched characteristics up to n^*, but not beyond that. This restriction in terms of the bidding language would then lead to the optimal bids as described by (15).

In the context of digital advertising, the information structure can be given a second interpretation. In the standard auction setting, we fix an object and then the value of the bidder for the object is random and given by a distribution F. An alternative interpretation is that the valuation of the bidder, his taste, is fixed and the objects are drawn at random. A particular expected value for each bidder is then attained by bundling objects of different characteristics so that a given expected value is attained. In this interpretation, the seller has to offer the bidder a particular distribution of impressions to attain a particular expectation. Namely, all objects that offer more matches than the threshold level n^* are sold in package, whereas all objects with fewer than n^* matches are sold and priced in fine segments equal to the number of matches.

An important qualification is that our results suggest that in the presence of heterogeneous bidders, the conflation of objects is personalized, and optimally dependent on the preferences of each bidder. Thus, the items are *not* conflated uniformly across bidders, but in a manner that depends on the preferences elicited from the bidders.

6 Discussion and Conclusion

We presented a model of attributes of an object and preferences of the bidders that generated a model of independent private values. The seller controls the revenue in the second price auction through the elicitation of preferences and the flow of information into the bids. By disclosing only limited information, the seller can increase the revenue at the expense of some efficiency losses. The optimal information design can be interpreted as form of personalized conflation.

Manual Bidding and Obedience We discussed how the optimal information structure can be implemented by automated bidding algorithms. These algorithms generate bids for the advertisers as function of the preferences elicited from the advertisers and the attributes of the viewer. We established in Proposition 5 that automated bidding supports truthful revelation of the preference information by the advertisers. A different class of algorithms is frequently described as manual bidding algorithms. Here, the bidders are asked to reveal their preferences first, and then are invited to bid on the basis of bid recommendations that take into account both preference and attribute information. These algorithms implicitly require both truthtelling and obedience constraints to be satisfied. The additional restrictions imposed by the obedience constraints, namely that the bidder wishes to follow the bid recommendation, may sometimes prevent the implementation of the optimal information structure for a given number N of bidders. Yet, we can show that in the presence of a large number of bidders we can approximate the revenue of the optimal information structure even when we impose the dual incentive constraints of thruthtelling and obedience on the information design. Namely, even if the seller has to elicit the private taste vector of the bidders in an incentive compatible manner, the optimal revenue can be approximated arbitrarily closely. Thus, there is a class of information structures that can appropriately balance revenue-maximization and incentive compatibility when there is a large number of bidders.

Horizontal vs. Vertical Differentiation An interpretation of the characteristics model is that it generates horizontal differentiation among bidders. By contrast, other models of characteristics would generate vertical differentiation and correlated values. For example, if each dimension of attributes and tastes were represented by the positive orthant, then a higher realization of any attribute would represent an element of vertical differentiation, and the ex-ante distribution of values would be correlated across bidders, thus introducing an element of common values. If valuations are correlated among bidders, then bidders would be subject to the winner's curse. This would introduce new trade-offs that relate the bidders' information to the winner's curse. From a technical perspective, the second price would no longer have a dominant-strategy equilibrium, so the characterization of the revenue in terms of the order statistic of expected valuations would no longer hold. While extending the analysis formally is not trivial, the linkage principle suggests that the seller would publicly disclose any shocks about common shocks affecting correlations, thus making the valuations conditionally independent (which would allow us to go back to our analysis). Bergemann, Brooks, and Morris (2017), (2019) analyze in detail how correlated information across bidders can improve the revenue performance of standard auctions, such as first and second price auctions.

Auction Format In the current environment with independent private values, the revenue equivalence result holds. Thus all classic auction formats generate the same expected revenue. Hence, while we formally study the second-price auction, the results extend to all classic auction formats, e.g., first-price auction, all-pay auction as long as we maintain to sell the object with probability one.

Reserve Price Our main analysis focuses on the second-price auction without reserve price. However, the analysis of the optimal information structure can be extend to the auctions with a reserve price r with minor modifications. With a reserve price r, the optimal information structure now displays two pooling regions. Next to the pooling region at the top that remains, there is now a pooling region of value such that the expected valuation in this lower pool exactly matches the reserve price r. The details are spelt out in the Appendix. There remains an intermediate interval in which the bidders learn their valuation.

In the presence of the reserve price, the logic of the lower pooling region is similar to the pooling region at the top. Given that there is a reserve price r, which acts as if there is a competing bid, the information structure create a thicker market by pooling the valuations even if the pooling leads to a loss in efficiency.

Asymmetric Information Structures We focus on describing the optimal symmetric information structure. While we do not have a general result showing that the optimal information structure is always symmetric, we have some partial results that point to this direction. The symmetric information structure that we derive is indeed the unique optimal information structure when there are two or three bidders, thus $N = 2$ or $N = 3$. We can also show that, if the information structure is the optimal symmetric information for $N - 1$ bidders, then it is optimal for the remaining bidder to also observe the optimal symmetric information. Hence, there is no improvement that involves changing the information structure of only one bidder. The detailed results are in the Appendix.

References

Aggarwal, G., A. Badanidiyuru, and A. Mehta (2019): "Autobidding with Constraints," in *International Conference on Web and Internet Economics*, pp. 17–30. Springer.

Alonso, R., and O. Camara (2016): "Persuading Voters," *American Economic Review*, 106, 3590–3605.

Arnosti, N., M. Beck, and P. Milgrom (2016): "Adverse Selection and Auction Design for Internet Display Advertising," *American Economic Review*, 106, 2852–2866.

Bergemann, D., B. Brooks, and S. Morris (2017): "First Price Auctions with General Information Structures: Implications for Bidding and Revenue," *Econometrica*, 85, 107–143.

——— (2019): "Revenue Guarantee Equivalence," *American Economic Review*, 109, 1911–1929.

Bergemann, D., and M. Pesendorfer (2007): "Information Structures in Optimal Auctions," *Journal of Economic Theory*, 137, 580–609.

Blackwell, D. (1951): "Comparison of Experiments," in *Proc. Second Berkeley Symp. Math. Statist. Probab.*, pp. 93–102. University of California Press, Berkeley.

Board, S. (2009): "Revealing Information in Auctions: The Allocation Effect," *Economic Theory*, 38, 125–135.

Cohen, M., I. Lobel, and R. P. Leme (2020): "Feature-Based Dynamic Pricing," *Management Science*, 66, 4921–4943.

Dar, E. E., V. Mirrokni, S. Muthukrishnan, Y. Mansour, and U. Nadav (2009): "Bid Optimization for Broad Match Auctions," *WWW 2009 Proceedings of the 18th International Conference on World Wide Web.*

Deng, Y., J. Mao, V. Mirrokni, and S. Zuo (2021): "Towards Efficient Auctions in an Auto-Bidding World," in *Proceedings of the International World Wide Web Conference.*

Dworczak, P., and G. Martini (2019): "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, 127, 1993–2048.

Eliaz, K., and R. Spiegler (2016): "Search Design and Broad Matching," *American Economic Review*, 106, 563–586.

Facebook Business Help Center (2021): "About Automated App Ads," *https://www.facebook.com/business/help/309994246788275?id=1858550721111595, Accessed 16th May 2021.*

Ganuza, J.-J. (2004): "Ignorance Promotes Competition: An Auction Model of Endogeneous Private Valuation," *RAND Journal of Economics*, 35.

Google Ads Help Center (2021a): "About App campaigns," *https://support.google.com/google-ads/answer/6247380?hl=en, accessed May 25th, 2021.*

——— (2021b): "About Automated Bidding," *https://support.google.com/google-ads/answer/2979071?hl=en, Accessed 15 May 2021.*

——— (2021c): "Broad Match," *https://support.google.com/google-ads/answer/2407779?hl=en*, accessed May 31st 2021.

Hartline, J., A. Johnsen, D. Nekipelov, and O. Zoeter (2019): "Dashboard Mechanisms for Online Marketplaces," in *Proceedings of the 20th ACM Conference on Economics and Computation.*

Kleiner, A., B. Moldovanu, and P. Strack (2021): "Extreme Points and Majorization: Economic Applications," *Econometrica*, forthcoming‘.

Kolotilin, A., T. Mylovanov, and A. Zapechelnyuk (2021): "Censorship as Optimal Persuasion," *Theoretical Economics*, forthcoming.

Levin, J., and P. Milgrom (2010): "Online Advertising: Heterogeneity and Conflation in Market Design," *American Economic Review Papers and Proceedings*, 100, 603–607.

Palfrey, T. (1983): "Bundling Decisions by a Multiproduct Monopolist with Incomplete Information," *Econometrica*, 51, 463–483.

Shaked, M., and J. Shanthikumar (2007): *Stochastic Orders*. Springer Verlag.

Sorokin, C., and E. Winter (2021): "Pure Information Design in Classic Auctions," .

7 Appendix

Proof of Proposition 4. Before we begin, it is useful to establish the rate at which ρ as defined earlier in (9) converges to 1 as N diverges to infinity. In the limit, the optimal quantile satisfies:

$$\lim_{N\to\infty} \frac{1-q_N^*}{1/N} = \rho,$$

with ρ satisfying (8).[5] Hence, for N large enough, $q \approx 1-\rho/N$, for some $\rho \in \mathbb{R}$

In what follows, for any two positive functions $H, \widetilde{H}$,

$$H(t) \sim \widetilde{H}(t), \text{ as } t \to \infty$$

means that

$$\lim_{t\to\infty} \frac{H(t)}{\widetilde{H}(t)} = 1$$

Throughout the proof, we use the following results about regularly-varying functions. First, for any regularly varying function $H(t)$ with index γ, there exists a slowly varying function $l(t)$ such that:

$$H(t) = l(t)t^{\gamma}.$$

Second, the slowly varying function behaves as a constant under integration of the tail:

$$\int_t^{\infty} l(y)y^{\gamma}dy \sim -l(t)t^{\gamma+1}(\gamma+1)^{-1}, \text{ as } t \to \infty$$

whenever $\gamma < -1$.

We can then write the density as follows:

$$f(t) = -\alpha l(t)t^{\alpha-1}.$$

In the limit $t \to \infty$, the hazard rate satisfies:

$$\bar{F}(t) \triangleq 1 - F(t) \sim l(t)t^{\alpha}, \text{ as } t \to \infty.$$

We thus have that $\bar{F}$ is also a slowly varying function.

[5] We can verify this claim, by noting that if we replace $q = 1-\rho/N$ in equation (5) and take the limit, we get this expression for ρ. Similarly, if q converges to 1 at a faster or slower rate than $1/N$, then clearly (5) cannot be satisfied in the limit.

We consider the following upper-pooling information structures:

$$G^{-1}(q) = \begin{cases} F^{-1}(q) & q \leq 1 - \frac{\rho}{N}; \\ \frac{\int_{F^{-1}(1-\frac{\rho}{N})}^{\infty} t dF(t)}{1/N} & \text{if } q > 1 - \frac{\rho}{N}, \end{cases} \tag{16}$$

with $\rho \in \mathbb{R}$. The quantile threshold is not necessarily the same as in the optimal information structure. We define:

$$V(N) \triangleq F^{-1}(1 - \frac{\rho}{N}),$$

which is the value at which the pooling zone begins. We denote by R' the expected revenue generated by this information structure.

We denote the difference between revenue generated under information structure (16) and under complete information as follows:

$$\Delta R \triangleq R' - R_c.$$

Since $w_{(2)} = v_{(2)}$ whenever $v_{(2)} < V(N)$, we can write the difference as follows:

$$\begin{aligned} \Delta R =& \mathbb{P}(v_{(2)} \geq V(N))\mathbb{E}[v_i \mid v_i \geq V(N)] - \mathbb{E}[v_{(2)} 1_{v_{(2)} \geq V(N)}] \\ =& \frac{\int_{V(N)}^{\infty} t dF(t)}{\rho/N}(1 - S(1-\rho/N)) - \int_{V(N)}^{\infty} tN(N-1)F^{N-2}(t)(1-F(t))dF(t). \end{aligned}$$

Finally, we can bound the difference as follows:

$$\Delta \tilde{R} \triangleq \frac{\int_{V(N)}^{\infty} t dF(t)}{\rho/N}(1 - S(1-\rho/N)) - \int_{V(N)}^{\infty} tN(N-1)(1-F(t))dF(t) \leq \Delta R,$$

where we omitted the term F^{N-2} in the integral and denoted the lower bound by $\Delta \tilde{R}$. Finally, we give a bound on the revenue generated under complete information:

$$\begin{aligned} \tilde{R}_c \triangleq& S(1-\rho/N)V(N) + \int_{V(N)}^{\infty} tN(N-1)(1-F(t))f(t)dt \\ \geq& \int_{0}^{V(N)} tN(N-1)(1-F(t))F^{N-2}(t)f(t)dt + \int_{V(N)}^{\infty} tN(N-1)(1-F(t))F^{N-2}(t)f(t)dt = \mathbb{E}[v_{(2)}] \end{aligned}$$

where we obtained the upper bound by omitting the term F^{N-2} in the integral, replaced t in the integral with $V(N)$, and denoted the lower bound by $\tilde{R}_c$. We note that:

$$\frac{R}{R_c} - 1 \geq \frac{\Delta R}{R_c} \geq \frac{\Delta \tilde{R}}{\tilde{R}_c}.$$

We prove that the right-hand-side of the inequalities does not converge to 0 as N grow large.

In the limit $N \to \infty$,

$$\int_{V(N)}^{\infty} tdF(t) \sim \alpha \frac{l(V(N))(V(N))^{(\alpha+1)}}{(\alpha+1)}$$
$$\int_{V(N)}^{\infty} tN(N-1)(1-F(t))dF(t) \sim \alpha \frac{N(N-1)}{(2\alpha+1)} l(V(N))^2 V(N)^{2\alpha+1}$$
$$(1 - S(1-\rho/N)) \sim \frac{e^{\rho} - \rho - 1}{e^{\rho}}$$

We now recall that:

$$l(V(N))V(N)^{\alpha} \sim \bar{F}(1-\rho/N) = \frac{\rho}{N}, \text{ as } N \to \infty. \tag{17}$$

We thus have the following approximations:

$$\int_{V(N)}^{\infty} tdF(t) \sim \alpha\rho \frac{V(N)}{N(\alpha+1)};$$
$$\int_{V(N)}^{\infty} tN(N-1)(1-F(t))dF(t) \sim \alpha \frac{N(N-1)}{(2\alpha+1)} \frac{\rho^2}{N^2} V(N).$$

So we conclude that:

$$\Delta\tilde{R} \sim V(N)\left(\frac{\alpha}{(\alpha+1)} \frac{e^{\rho} - \rho - 1}{e^{\rho}} - \frac{\alpha\rho^2}{(2\alpha+1)}\right), \text{ as } N \to \infty.$$

Using the same calculations as before, in the limit $N \to \infty$:

$$\tilde{R}_c \sim V(N)\left(\frac{1+\rho}{e^{\rho}} + \alpha \frac{1}{(2\alpha+1)}\right)$$

So, we have that:

$$\frac{\Delta\tilde{R}}{\tilde{R}_c} \sim \frac{\left(\frac{\alpha}{(\alpha+1)} \frac{e^{\rho}-\rho-1}{e^{\rho}} - \frac{\alpha\rho^2}{(2\alpha+1)}\right)}{\left(\frac{1+\rho}{e^{\rho}} + \alpha \frac{1}{(2\alpha+1)}\right)}, \text{ as } N \to \infty$$

Finally, in the limit $\rho \to 0$,

$$\lim_{\rho \to 0} \frac{\frac{e^{\rho}-\rho-1}{e^{\rho}}}{\rho^2} = \frac{1}{2}.$$

However, we also have that:

$$\left(\frac{\alpha}{(\alpha+1)} \frac{1}{2} - \frac{\alpha}{(2\alpha+1)}\right) > 0.$$

Hence, $\alpha\left(\frac{1}{(\alpha+1)} \frac{e^{\rho}-\rho-1}{e^{\rho}} - \frac{\rho^2}{(2\alpha+1)}\right) > 0$ for a small enough ρ, so get that:

$$\lim_{N \to \infty} \frac{R}{R_c} - 1 > 0.$$

This proves that the ratio (11) converges to a number larger than 1. Furthermore, in the limit $\alpha \to -1$, we have that:

$$\lim_{\alpha\to-1}\lim_{N\to\infty}\frac{R}{R_c}-1=\infty.$$

This limit holds uniformly for every ρ. This proves that the ratio diverges in the limit $\alpha \to -1$.

Finally, we prove that z in (11) exists (i.e., it is not infinite). For this, we now define:

$$\begin{aligned}\Delta\tilde{R}' \triangleq& \frac{\int_{V(N)}^{\infty} t dF(t)}{1/N} S(1-1/N) - F^{N-2}(V(N))\int_{V(N)}^{\infty} tN(N-1)(1-F(t))dF(t)\\ &\tilde{R}_c'^{N-2}(V(N))\int_{V(N)}^{\infty} Nt(1-F(t))f(t)dt.\end{aligned}$$

and note that:

$$\frac{R}{R_c}-1\le\frac{\Delta\tilde{R}'}{\tilde{R}_c'}.$$

The difference with the lower bounds previously calculated is that instead of omitting the term F^{N-2}, we evaluate it at the lower limit of the integral. Following similar steps as before, we have that, as $N\to\infty$:

$$\begin{aligned}\Delta\tilde{R}' \sim& -\alpha V(N)\left(\frac{1}{(\alpha+1)}\frac{e^{\rho}-\rho-1}{e^{\rho}}-\frac{\rho^2}{e^{\rho}}\frac{1}{(2\alpha+1)}\right),\\ \tilde{R}_c' \sim& -\frac{1}{e^{\rho}}\alpha\frac{1}{(2\alpha+1)}V(N).\end{aligned}$$

We thus conclude that, in the limit $N\to\infty$:

$$\frac{R}{R_c}-1\le\frac{\Delta\tilde{R}'}{\tilde{R}_c'}\sim\frac{-\alpha\left(\frac{1}{(\alpha+1)}\frac{e^{\rho}-1-\rho}{e^{\rho}}-\frac{\rho^2}{e^{\rho}}\frac{1}{(2\alpha+1)}\right)}{\frac{1}{e^{\rho}}\alpha\frac{1}{(2\alpha+1)}}<\infty,$$

for all $\rho\in\mathbb{R}$. Finally, we note that the optimal quantile converges to 1 at a rate of ρ/N, so the upper pooling information structure we are studying converges to the optimal one for some ρ. This proves that the limit z does not diverge. ■

8 Appendix with Extensions

8.1 Manual Bidding

We now examine a model of manual bidding. We suppose that the seller needs to elicit the preference y_i and then present the bidder a signal as a function of the realized match quality (based on the reported preference). The buyer can then freely choose his bid based on the information presented to him. As before, the buyers bid in a second-price auction without reserve, and the seller's objective is to maximize revenue. While the optimal information structure may not be incentive compatible, we provide another information structure that is incentive compatible and generates a revenue approximately optimal when the number of bidders is large.

8.1.1 Model of Manual Bidding

To analyze the incentives of bidders to truthfully report their preferences to the publisher, we need to make further assumptions about the payoff environment. In particular, we will assume that the valuation function is as in (13). As before, attributes and preferences take values in $\{-1, 1\}$, are uniformly and independently distributed across components and bidders.

As with autobidding, a reporting strategy for bidder i is denoted by:

$$\widetilde{y}_i : \{-1, 1\}^J \to \Delta\{-1, 1\}^J.$$

Given the reported preferences, the seller discloses to the bidder a signal $s(\widetilde{v}_i)$, where

$$\widetilde{v}_i \triangleq u(\frac{1}{\sqrt{J}} \sum_{j=1}^{J} \widetilde{y}_{ij}(y_{ij}) x_j).$$

That is, $\tilde{v}_i$ is the valuation of an advertiser with preferences $\tilde{y}_i$.

We denote by $\widehat{w}_i$ the expected value of v_i conditional on $s(\widetilde{v}_i)$:

$$\widehat{w}_i \triangleq \mathbb{E}[v_i \mid s(\widetilde{v}_i), y_i].$$

and by $\widehat{G}_i$ the distribution of expected valuations. As before, w_i denotes the expected valuation when a bidder reports truthfully, and G_i denotes the respective distribution. The seller's problem is then to find a information structure that solves:

$$R \triangleq \max_{\{s:\mathbb{R}\to\Delta\mathbb{R}\}} \mathbb{E}[w_{(2)}]; \tag{18}$$

$$\mathbb{E}[\max\left\{w_i - \max\{w_j\}_{j\neq i}\,,\, 0\right\}] \geq \mathbb{E}[\max\left\{\widehat{w}_i - \max\{w_j\}_{j\neq i}\,,\, 0\right\}],\ \text{for all } \widetilde{y}_i \tag{19}$$

Here the constraint is the incentive compatibility constraint: the expected bidder surplus is weakly larger when reporting truthfully than any other reporting strategy. Here we require Bayesian incentive compatibility to keep the notation more compact, but this will play no role.

8.1.2 Incentive Compatibility Under Manual Bidding

Since the preferences and attributes are symmetrically distributed, a sufficient statistic for the bidder's strategy is the fraction of preferences truthfully reported:

$$\rho_i \triangleq \sum_{j=1}^{J} \frac{\widetilde{y}_{ij} y_{ij}}{J}.$$

Note that ρ_i is the correlation between the reported preference and the real preference. In other words, for any reporting strategy $\widetilde{y}_i, \widetilde{y}_i'$ satisfying $\rho_i = \rho_i'$, the induced distribution of expected valuations will be the same. In this case, $\widetilde{y}_i$ and $\widetilde{y}_i'$ are equivalent from the perspective of the information generated for the bidder. If $\rho_i = 1$ then the preference has been correctly reported; if $\rho_i = 0$ then half of all preference components have been misreported; if $\rho_i = -1$ then every preference component has been incorrectly reported.

Since $\tilde{v}_i$ is a a noise signal about v_i, a natural conjecture is that bidders will want to report their preference truthfully. However, misreporting *every* preference component (i.e., $\rho_i = -1$) may sometimes be profitable. In this case, the bidder will observe signal $s(u(-m_i))$ (instead of $s(u(m_i))$). In this case, the distribution of expected valuations will be:

$$\widehat{G}^{-1}(t) \triangleq \begin{cases} F^{-1}(t) & \text{for all } t \geq 1 - q_N^* \\ \frac{\int_0^{1-q_N^*} F^{-1}(q)dq}{q_N^*} & \text{for all } t < 1 - q_N^* \end{cases}$$

with q_N^* defined in (5). This expression is akin to (6), but the pooling section is at the lower quantile instead of the upper quantile.

We begin by establishing that the only relevant incentive constraints are those induced by reporting the exact opposite preference.

Lemma 1 (Informativeness of Signals)
Let s be the optimal information structure. The generated signal for every $\rho_i \in [0, 1)$ is less informative than the signal generated for $\rho_i = 1$. The generated signal for every $\rho_i \in [-1, 0)$ is less informative than the signal generated for $\rho_i = -1$.

Proof of Lemma 1. We prove the statement for $\rho_i \in [0,1]$. We denote by $\widehat{F}$ the distribution of the expected value of v_i conditional on $\widetilde{v}_i$ (i.e., $\mathbb{E}[v_i \mid \widetilde{v}_i] \sim \widehat{F}$). We denote by $\widehat{G}$ the distribution of $\widehat{w}_i$, which is given by:

$$\widehat{G}^{-1}(t) = \begin{cases} \widehat{F}^{-1}(t) & \text{for all } t < q_N^* \\ \frac{\int_{q_N^*}^1 \widehat{F}^{-1}(q)dq}{1-q_N^*} & \text{for all } t \in [q_N^*, 1) \end{cases}$$

with q_N^* defined in (5) (this is simply (6) but replacing F with $\widehat{F}$).

We first observe that $\widetilde{v}_i$ is a noisy signal of v_i. Hence, $\widehat{F}$ is a mean preserving contraction of F, which is equivalent to stating that:

$$\int_0^t F^{-1}(q)dq \leq \int_0^t \widehat{F}^{-1}(q)dq$$

for all t with equality for $t = 1$. We thus have that:

$$\int_0^t G^{-1}(q)dq \leq \int_0^t \widehat{G}^{-1}(q)dq$$

for all $t \leq q_N^*$ (in this range $\widehat{G}^{-1}(t) = \widehat{F}^{-1}(t)$ and $G^{-1}(t) = F^{-1}(t)$). Since

$$\int_0^1 \widehat{G}^{-1}(q)dq = \int_0^1 G^{-1}(q)dq$$

and $\widehat{G}^{-1}(t)$, $G^{-1}(t)$ are constant for $t > q_N^*$ we must have that:

$$\frac{\int_{q_N^*}^1 \widehat{F}^{-1}(q)dq}{1-q_N^*} \leq \frac{\int_{q_N^*}^1 F^{-1}(q)dq}{1-q_N^*}$$

and

$$\int_0^t \widehat{G}^{-1}(q)dq \leq \int_0^t G^{-1}(q)dq$$

for all $t \geq q_N^*$, with equality only for $t = 1$. This proves the result for $\rho_i \in [0,1]$.

The case $\rho_i \in [-1,0]$ can be proved in a completely analogous way except for the fact that the distribution of expected valuations under signals $s(-m_i)$ is:

$$\widehat{G}^{-1}(t) = \begin{cases} \widehat{F}^{-1}(t) & \text{for all } t \geq 1 - q_N^* \\ \frac{\int_0^{q_N^*} \widehat{F}^{-1}(q)dq}{q_N^*} & \text{for all } t < 1 - q_N^* \end{cases}$$

and analogously for $s(\widetilde{v}_i)$. This is because the signals generate pooling for low quantiles instead of high quantiles. The rest of the proof proceed in a completely analogous way. ■

The lemma shows that there are two strategies that lead to the most informative signals: reporting truthfully and reporting the opposite preference.

It is possible to verify that reporting truthfully may fail to be an equilibrium for some model parameters (in particular, for some N, u).[6] The reason is that by misreporting his preference, an advertiser may gain better information about his own valuation when this is high. This is because when he misreports his preferences the distribution of expected valuations is fully informative at the top but has a pooling section at the bottom. Depending on the shape of u and the number of advertisers, reporting truthfully may fail to be an equilibrium (in fact, it will not be an equilibrium unless u is sufficiently concave).

8.1.3 Approximately Optimal Mechanism

Our analysis here pursues a limited objective. We do not attempt to characterize the set of feasible and incentive compatible mechanism which would be an interesting avenue to pursue. Rather, we suggest a small modification of the optimal mechanism of Proposition 1 and show that the modified mechanism is indeed incentive compatible. Moreover, as the number of bidders becomes large, the revenue of the modified mechanism approximates the mechanism of the optimal information structure.

While truthtelling will not be an equilibrium for every $N,\ u$ there is a class of information structures that can appropriately balance revenue-maximization and incentive compatibility when there is a large number of bidders. Consider the following information structure.

$$s(v_i) = \begin{cases} \mathbb{E}[v_j \mid F(v_j) \leq 1-q] & \text{if } F(v_j) \leq 1-q_N^* \\ v_j & \text{if } 1-q_N^* \leq F(v_j) \leq q_N^* \\ \mathbb{E}[v_j \mid F(v_j) \geq q] & \text{if } F(v_j) \geq q_N^* \end{cases} \tag{20}$$

with q_N^* defined in (5). In other words, the information structure is as the optimal information structure (4) but in addition to the pooling at the top, there is pooling at the bottom.

Proposition 6 (Incentive Compatibility)

Under manual bidding and information structure (20)*, it is a dominant strategy for the advertiser to report his preference truthfully.*

Proof of Proposition 6. It is easy to check that Lemma 1 also holds for the two-sided pooling information structures. We now notice that $s(v_i) = s(-v_i)$, so reporting truthfully or the exact opposite

[6] For example, if F has a mass of size q_N^* at $v = 0$, the optimal information structure will not be incentive compatible. This can be immediately verified by noting that reporting the exact opposite preference (i.e., $\rho_i = -1$) allows the bidder to learn his value perfectly.

generates the same signal. Hence, reporting truthfully generates the most informative signal about an bidder's own type, so this is a dominant strategy. ■

Under the optimal information structure, truthtelling might not be an equilibrium because bidders have an incentive to report the exact opposite of their preferences. Because the two-sided pooling information is symmetric – that is, there is the same pooling zone for high and low quantiles – the incentive to misreport the tastes disappears. In particular, reporting truthfully and reporting the exact opposite preference generates the same signal for the bidder.

Information structure (20) will not maximize revenue across all information structure, but the loses from having pooling at the bottom as the number bidders becomes large is negligible. This is because the probability that the valuation of the second-highest bidder is smaller than $F^{-1}(1-q_N^*)$ converges to zero as N becomes large. Note that there are two complementary effects by which this probability converges to 0. First, $(1-q_N^*)$ converges to 0 as N becomes large, so the probability that the valuation of any bidder is below this threshold converges to 0. Second, for any fixed quantile q the probability that the valuation of the second-highest bidder is smaller than $F^{-1}(1-q_N^*)$ converges to zero as N becomes large. Hence, the revenue losses from having pooling at the bottom vanish due to both effects, and are expected to be small.

Proposition 7 (Approximate Optimality)
Under the two-sided pooling information structure the revenue converges to the one under the optimal information structure when the number of bidders grows large:

$$\lim_{N\to\infty}(\mathbb{E}[w_{(2)}]-R)=0$$

Proof of Proposition 7. We denote by H (resp. G) the distribution of expected valuations induced by the two-sided pooling information structure (resp. optimal information structure) and write the difference as follows:

$$(\mathbb{E}[w_{(2)}]-R)=\int_0^1 H^{-1}(q)S'(q)dq-\int_0^1 G^{-1}(q)S'(q)dq$$

Whenever $v_i \geq G^{-1}(1-q)$ the two-sided pooling information structure and the optimal information structures are the same, so $H^{-1}(q)=F^{-1}(q)$. We thus have that,

$$(\mathbb{E}[w_{(2)}]-R)=\int_0^{1-q}(H^{-1}(q)-G^{-1}(q))S'(q)dq.$$

In the limit $N\to\infty$, we have that $q\to 0$ and $S'(q)\to 0$ for every fixed q. Since $H^{-1}(q)$ and $G^{-1}(q)$ are bounded in the interval $[0,1-q]$, we must have that

$$\lim_{N\to\infty}(\mathbb{E}[w_{(2)}]-R)=0$$

which establishes the result. ■

The proposition states that the two-sided pooling information structure can effectively balance the trade-off between incentive compatibility and revenue maximization when there is a large number of bidders. This result is specially relevant when the distribution of valuations has a thick tail because in this case the gains from using the optimal information structure (instead of complete information) do not vanish. Finding the optimal incentive compatible information structure is an interesting open question that is left for future work.

8.2 Reserve Price

We now assume that the second-price auction has reserve price $r > 0$. As before, we can assume that the bid is equal to an advertiser's expected valuation w_i.

Proposition 8 (Optimal Information Structure)
There are quantiles q_1, q_2, q_3 such that an optimal information structure is given by:

$$s(v_i) = \begin{cases} s_1 & \text{if } F(v_i) \in [0, q_1) \\ s_2 & \text{if } F(v_i) \in [q_1, q_2) \\ v_i & \text{if } F(v_i) \in [q_2, q_3] \\ s_3 & \text{if } F(v_i) \in [q_3, 1] \end{cases} \tag{21}$$

The intervals may be degenerate, and $\mathbb{E}[v_i \mid s = s_2] = r$.

Proof. The expected revenue is given by:

$$R = \mathbb{P}\{w_{(1)} \geq r \text{ and } w_{(2)} < r\}r + \mathbb{E}[w_{(2)} 1_{w_{(2)} \geq r}].$$

For any distribution of expected valuations G, the distribution of $w_{(2)}$ is given by $(nG^{n-1}(x)(1 - G(x)) + G^n(x))$ and

$$\mathbb{P}\{w_{(1)} \geq r \text{ and } w_{(2)} \leq r\} = nG^{n-1}(r)(1 - G(r)),$$

so we can write the revenue as follows:

$$R = 1 - rG(r)^n - \int_r^1 (nG^{n-1}(x)(1 - G(x)) + G^n(x))dx.$$

Let $\hat{G}$ be the distribution of expected valuations induced by the optimal information structure. Let $\hat{q}_r = \hat{G}(r)$ and $\hat{Q} = \hat{G}^{-1}$. Then, we can write it as follows

$$R = 1 - \hat{q}_r^n \hat{Q}(\hat{q}_r) - \int_{\hat{q}_r}^{1} (nq^{n-1}(1-q) - q^n) d\hat{Q}(q)$$

The majorization constraint states that for all $x \in [0,1]$:

$$\int_x^1 \hat{Q}(q)dq \leq \int_x^1 F^{-1}(q)dq.$$

Let $\psi \in [0,1]$ be such that,

$$\int_{\hat{q}_r}^1 \hat{Q}(q)dq = \int_{\hat{q}_r}^1 \mathbf{1}_{\{q\geq\psi\}} F^{-1}(q) + \mathbf{1}_{\{q<\psi\}} \hat{Q}(\hat{q}_r)dq \tag{22}$$

Let $\tilde{F}^{-1} : [\hat{q}_r, 1] \to [0,1]$ be defined as follows:

$$\tilde{F}^{-1}(q) = \mathbf{1}_{\{q\geq\psi\}} F^{-1}(q) + \mathbf{1}_{\{q<\psi\}} \hat{Q}(\hat{q}_r)$$

Consider the following maximization problem:

$$Q_r \triangleq \underset{Q:[\hat{q}_r,1]\to[\hat{Q}(\hat{q}_r),1]}{\arg\max} \; - \int_{\hat{q}_r}^1 (nq^{n-1}(1-q) - q^n)dQ(q) \tag{23}$$

$$\text{subject to: for all } q \geq \hat{q}_r, \; \int_x^1 Q(q)dq \leq \int_x^1 \tilde{F}^{-1}(q)dq \text{ with equality when } x = \hat{q}_r, \tag{24}$$

We first note that $\hat{Q}(x) = Q_r(x)$. To verify this, consider the following function:

$$\hat{Q}'(q) = \begin{cases} \hat{Q}(q) & \text{if } q \leq \hat{q}_r \\ Q_r(q) & \text{otherwise} \end{cases}$$

The revenue under $\hat{Q}'$ is given by:

$$\begin{aligned} R =& 1 - \hat{q}_r \hat{Q}'(\hat{q}_r) - \int_{\hat{q}_r}^1 (nq^{n-1}(1-q) - q^n)d\hat{Q}'(q) \\ =& 1 - \hat{q}_r \hat{Q}(\hat{q}_r) - \int_{\hat{q}_r}^1 (nq^{n-1}(1-q) - q^n)dQ_r(q) \\ \geq& 1 - \hat{q}_r \hat{Q}(\hat{q}_r) - \int_{\hat{q}_r}^1 (nq^{n-1}(1-q) - q^n)d\hat{Q}(q) \end{aligned}$$

where the last term is the revenue under $\hat{Q}$. Hence, $\hat{Q}'$ is optimal. We also note that:

$$\int_x^1 \hat{Q}'(q)dq \leq \int_x^1 \hat{Q}(q)dq,$$

with equality for $x \leq \hat{q}_r$. Hence, $\hat{Q}'$ satisfies the majorization constraint. The inequality follows from the fact that Q_r satisfies (24) and ψ is such that (22) is satisfied.

We now note that (23) is the same problem that we solved to prove Propositions 1 but restricted to $[\hat{q}_r, 1]$. Hence, $Q_r(q)$ restricted to $[\hat{q}_r, 1]$ is

$$Q_r(q) = \begin{cases} \tilde{F}^{-1}(q) & \text{if } q \in [\hat{q}_r, q_N^*] \\ \mathbb{E}[v_j \mid F(v_j) \geq q] & q \in [q_N^*, 1) \\ 1 & q = 1 \end{cases}$$

for some q_N^*. By definition, $\hat{Q}(q) < r$ for all $q < \hat{q}_r$, so the expected valuation induced by these quantiles is less than the reserve price. Hence, the distribution of expected valuations below this quantile is irrelevant. Finally, we note that the following information structure generates this distribution of expected valuations:

$$s(v_i) = \begin{cases} s_1 & \text{if } F(v_i) \in [0, \hat{q}_r) \\ s_2 & \text{if } F(v_i) \in [\hat{q}_r, \psi) \\ v_i & \text{if } F(v_i) \in [\psi, q^*] \\ s_3 & \text{if } F(v_i) \in [q^*, 1] \end{cases}$$

It is straightforward that information structure (21) generates the distribution of expected valuations (where we just replaced the specific quantiles with generic variables q_1, q_2, q_3). ■

The proposition shows that the information structure has three pooling intervals and one interval of full disclosure. The first interval consists of bidders who know their valuation is below r so they do not buy the good (i.e., interval $[0, q_1)$). The second interval consists of the bidders whose conditional expected valuation is r, so they buy the good. The third interval are the bidders who know their valuation. The fourth interval is the bidders who know their valuation is the highest possible. The last two intervals are the same as in the case without reserve price.

8.3 Asymmetric Information Structures

We now relax the assumption that the publisher is restricted to symmetric signals. That is, we allow for the possibility that $s_i \neq s_j$.

8.3.1 No Optimal Asymmetric Information Structure When $N = 2, 3$

We now show that there is no optimal information structure when $N = 2, 3$, which we state in the following proposition.

Proposition 9 (Uniqueness of the Optimal Information Structure)
If $N = 2$ or $N = 3$, then the optimal symmetric information structure is the unique optimal information structure.

The case $N = 2$ is straightforward to argue. In this case the revenue is the minimum between the expected valuation of the two bidders. Thus, giving no information maximizes revenue. A formal proof can be found in Board (2009). We thus focus on the case $N = 3$.

The distribution of the second-highest expected valuation is:

$$P(w_{(2)} \leq x) = G_1(x)G_2(x) + G_2(x)G_3(x) + G_1(x)G_3(x) - 2G_1(x)G_2(x)G_3(x)$$

Let $\Psi(F)$ be defined as follows:

$$\Psi \triangleq \{G : \; F \prec G \text{ and } G \text{ is monotonic}\}.$$

Integrating by parts, we can write (1) as follows:

$$\max_{G_1, G_2, G_3 \in \Psi} 1 - \int G_1(x)G_2(x) + G_2(x)G_3(x) + G_1(x)G_3(x) - 2G_1(x)G_2(x)G_3(x)dx \tag{25}$$

The objective function is linear on each of the functions G_1, G_2, G_3, but it is not jointly linear. This means that, holding fixed two of the distributions, say G_1, G_2, the maximization over G_3 is a classic problem of Bayesian persuasion with a continuous state space as studied by Dworczak and Martini (2019), among others. However, because the problem is not jointly linear, it is not possible to maximize over each of the functions independently. Furthermore, for some arbitrary functions G_1, G_2 the maximization problem over G_3 can attain its maximum over functions that do not resemble qualitatively those describe in Proposition 1. Note that because the problem is not linear, it is not even clear that the optimum will be attained at extreme points of Ψ.

We first argue that, if an asymmetric optimal information structure exists, then there also exists an optimal information structure in which G_1, G_2 and G_3 are extreme points of Ψ. For simplicity, we make the argument assuming that $G_2 = G_3 \in \Psi$, while G_1 is not. Since (25) is linear in G_1 and G_1 is in the interior of Ψ, we can find $G_1' \neq G_1''$ that are: (a) extreme points of Ψ, (b) in the support of G_1, and (c) the information structures (G_1', G_2, G_3) and (G_1'', G_2, G_3) yield the same expected revenue as (G_1, G_2, G_3). We can then without loss of generality find an optimal information structure (G_1', G_2, G_3) that is an optimal information structure, in which all elements are extreme points of Ψ and and such that $G_1' \neq G_2 = G_3$ (if $G' = G_2$, then we just pick G''). The case in which G_2 or G_3 are not extreme points of Ψ can be argued analogously.

We denote by β_j^1 the maximum in the support of G_j:

$$\beta_j^1 = \max\{x \in suppG_j\}. \tag{26}$$

Throughout the proof, we label agents such that $\beta_3^1 \leq \beta_2^1 \leq \beta_1^1$. Fixing the distribution G_1, G_2 and maximize over all distributions for bidder 3, we solve:

$$\max_{G \in \Psi} 1 - \int G_1(x)G_2(x) + G_2(x)G(x) + G_1(x)G(x) - 2G_1(x)G_2(x)G(x)dx$$

We write the maximization problem as follows:

$$\max_{G \in \Psi} \int G(-G_2(x) - G_1(x) + 2G_1(x)G_2(x))dx + constants \tag{27}$$

where the term "constants" refers to terms that do not depend on G. Of course, one can write analogously the optimization over the distribution of expected valuations for bidder 1 and 2 keeping the other two bidders fixed. This optimization problem will arise several times throughout the proofs.

Lemma 2
For every optimal information structure, $\beta_2^1 = \beta_1^1 < 1$.

Proof of Lemma 2. Suppose that $\beta_2^1 < \beta_1^1$. That is, the highest element in the support of the distribution of G_1 is strictly larger than the highest element in the support of the distribution of G_2, G_3. We write the expected revenue as follows:[7]

$$R = \mathbb{P}(w_1 > \beta_2^1)\mathbb{E}[w_{(2)} \mid w_1 > \beta_2^1] + \mathbb{P}(w_1 \leq \beta_2^1)\mathbb{E}[w_{(2)} \mid w_1 \leq \beta_2^1].$$

We note that:

$$\mathbb{E}[w_{(2)} \mid w_1 > \beta_2^1] > \mathbb{E}[w_{(2)} \mid w_1 \leq \beta_2^1].$$

[7]Recall that the subindex refers to the agent number when it is without parenthesis and it refers to the respective order statistic when it is with a parenthesis. That is, w_2 is the expected valuation of agent 2 and $w_{(2)}$ is the second-order statistic of the expected valuations.

That is, the expected revenue is strictly higher when the realization of bidder 1's expected valuation is β_1^1. We also note that

$$\mathbb{E}[w_{(2)} \mid w_1 >= w'] = \mathbb{E}[w_{(2)} \mid w_1 = w'']$$

for all $w', w'' > \beta_2^1$. That is, the expected revenue is the same whenever th realization of bidder 1's expected valuation is higher than β_2^1 (this is because the revenue is the second-highest expected valuation). So, without loss of generality we can assume that the distribution of expected valuations G_1 has a mass point at β_1^1 and every other element in the support is weakly lower than β_2^1.

We now consider the following information structure that is constructed based on the signal that generated G_1:

$$s_1 = \begin{cases} \beta_1^1 - \epsilon & \text{if } w_1 = \beta_1^1 \text{ or if with probability } \delta \text{ if } w_1 < \beta_1^1 \\ w_1 & \text{with probability } (1-\delta) \text{ if } w_1 < \beta_1^1 \end{cases}$$

We take δ and ϵ small enough such that $\beta_2^1 < \beta_1^1 - \epsilon$ and such that:

$$\mathbb{E}[w_1 \mid s_1 = \beta_1^1 - \epsilon] = \beta_1^1 - \epsilon.$$

We can then write the revenue under this new information structure as follows:

$$R = \mathbb{P}(w_1' > \beta_2^1)\mathbb{E}[w_{(2)} \mid w_1' > \beta_2^1] + \mathbb{P}(w_1' \leq \beta_2^1)\mathbb{E}[w_{(2)}' \mid w_1' \leq \beta_2^1].$$

Using the same arguments as before, we have that:

$$\mathbb{E}[w_{(2)}' \mid w_1' \leq \beta_2^1] = \mathbb{E}[w_{(2)} \mid w_1 \leq \beta_2^1] \text{ and } \mathbb{E}[w_{(2)}' \mid w_1' > \beta_2^1] = \mathbb{E}[w_{(2)} \mid w_1 > \beta_2^1].$$

However, we now have that $\mathbb{P}(w_1' > \beta_2^1) > \mathbb{P}(w_1 > \beta_2^1)$ so this new information structure generates a higher expected revenue. We thus conclude that $\beta_1^1 = \beta_2^1$.

We now suppose that $\beta_1^1 = \beta_2^1 = 1$. We then must have that the distribution of expected valuations is absolutely continuous in some neighborhood $[1-\delta, 1]$. In this case, we consider the following information structure:

$$s_1 = \begin{cases} 1 - \epsilon & \text{if } w_1 \geq 1 - e; \\ w_1 & \text{if } w_1 < \beta_1^1, \end{cases}$$

where

$$1 - \epsilon = \mathbb{E}[w_1 \mid w_1 \geq 1 - \delta].$$

We rewrite (27) but for bidder 1:

$$\max_{G \in \Psi} \int G(-G_2(x) - G_3(x) + 2G_3(x)G_2(x))dx + constants. \tag{28}$$

We note that $(-G_2(x) - G_3(x) + 2G_3(x)G_2(x))$ must be increasing in a neighborhood $[1-\delta, 1]$ for a small enough δ.[8] Hence, the new information structure generates higher revenue, which proves that $\beta_1^1 = \beta_2^1 < 1$. ■

This lemma implies that there exists $x < 1$ such that $G_1(x) = G_2(x) = G_3(x) = 1$. Theorem 2 in Kleiner, Moldovanu, and Strack (2021) implies that for every G_j, there exists $\tilde{v}_j < 1$ such that:

1. G is increasing at most in two points in $[\tilde{v}_j, 1]$;
2. $G(\tilde{v}_j) = F(\tilde{v}_j)$ and $\int_{\tilde{v}_j}^1 G(x)dx = \int_{\tilde{v}_j}^1 F(x)dx$;
3. $\int_y^1 F(x)dx < \int_y^1 G(x)dx$ for all $x \in (\tilde{v}_j, 1)$.

We denote by $\beta_j^2 < \beta_j^1$ the two steps of G_j and by Δ_j^1 and Δ_j^2 the size of these two steps. If G_j has only one step in $[\tilde{v}_j, 1]$ we adopt the convention that the step is at β_j^1 (which is consistent with (26)).

Lemma 3
Suppose that $\Delta_1^1, \Delta_2^1 \geq 1/2$, then a solution to (28) is:

$$G^*(x) = \begin{cases} F(x) & x \leq \bar{x}; \\ F(\bar{x}) & x \in [\bar{x}, \beta_1^1); \\ 1 & \textit{otherwise.} \end{cases} \tag{29}$$

where $\bar{x}$ solves $\beta_1^1 = \mathbb{E}[v \mid v \in [\bar{x}, 1]]$.

Proof of Lemma 3. We first note that, Lemma 2 implies that every solution to (28) satisfies $G_3(\beta_1^1) = 1$. Hence, we can, without loss of generality, the maximization problem (28) as follows:

$$\max_G \int_0^{\beta_1^1} G(x)(-G_2(x) - G_1(x) + 2G_1(x)G_2(x))dx + \textit{constants} \tag{30}$$
$$\text{subject to: } F \prec G \text{ and } G(\beta_1^1) = 1 \tag{31}$$

[8] This is immediate to check at every point of differentiability:

$$h'(x) = (-g_2(x) - g_3(x) + 2g_3(x)G_2(x) + 2G_3(x)g_2(x)) > (-g_2(x) - g_3(x) + g_3(x) + g_2(x)) = 0,$$

whenever $G_2(x), G_3(x) > 1/2$. However, $G_2(x), G_3(x) > 1/2$ for every x in a neighborhood $[1-\delta, 1]$. Clearly, at points of non-differentiability it must also be increasing.

Hence, we write the problem as a maximization with a majorization constraint with range in $[0, \beta_1^1]$. We define:

$$\widetilde{\Psi} \triangleq \{G : G \in \Psi \text{ and } G(\beta_1^1) = 1\},$$

and note that $G^* \in \widetilde{\Psi}$ (where G^* is defined in (29)).

We now note that for every $G \in \widetilde{\Psi}$, $G^* \prec G$. To verify this it is sufficient to check that $G^*(x) = F(x)$ for all $x \leq \bar{x}$ and by construction for all $G \in \Psi$, $\int_0^y G(x)dx \leq \int_0^y F(x)dx = \int_0^y G^*(x)dx$. Finally, since $G^*(x) = G(x)$ for all $x > \bar{x}$, we must have that for all $G \in \Psi$, $\int_0^y G(x)dx \leq \int_0^y G^*(x)dx$ for all $y \in [0, 1]$.

We now note that, for every $x < \beta_1^1$, $G_1(x), G_2(x) \leq 1/2$, and so the coefficient:

$$h(x) \triangleq (-G_2(x) - G_1(x) + 2G_1(x)G_2(x))$$

is decreasing in x.[9] Hence, G^* is an optimal solution, which follows from the Fan-Lorentz inequality (see Section 3.2.2. in Kleiner, Moldovanu, and Strack (2021)). ■

Lemma 3 shows that, when $\Delta_1^1, \Delta_2^1 \geq 1/2$ we will have that $\beta_3^1 = \beta_2^1 = \beta_1^1$. Furthermore, we have that $\Delta_3^1 = 1 - F(\bar{x})$, so it must also be that $\Delta_3^1 \geq \Delta_3^1, \Delta_2^1 \geq 1/2$. Using the same argument for bidder 1 and 2, we must have that $\Delta_3^1 = \Delta_2^1 = \Delta_1^1 = 1 - F(\bar{x})$ and the optimal information structure is symmetric. Since there is a unique symmetric information structure, which proves the result. We are thus left with proving that $\Delta_1^1, \Delta_2^1 \geq 1/2$. We state this formally in the following lemma and then prove it, which concludes the proof.

Lemma 4

In every optimal information structure $\Delta_1^1, \Delta_2^1 \geq 1/2$.

Proof. We establish the result by addressing separately the case $\beta_1^1 = \beta_2^1 > \beta_3^1$ and the case $\beta_1^1 = \beta_2^1 = \beta_3^1$, which we refer to as "Case 1" and "Case 2", respectively.

Case 1 We now assume that $\beta_1^1 = \beta_2^2 > \beta_3^1$ and show that a information structure is optimal only if $\Delta_1^1, \Delta_2^1 > 1/2$.

Sub-case A We begin by considering the case in which G_j has two steps in $[\tilde{v}_j, 1]$ for at least one of $j \in \{1, 2\}$. When G_j have two steps, then β_j^2 and Δ_j^2 are defined without ambiguity. If G_j has only one

[9] This is immediate to check at every point of differentiability:

$$h'(x) = (-g_2(x) - g_1(x) + 2g_1(x)G_2(x) + 2G_1(x)g_2(x)) < (-g_2(x) - g_1(x) + g_1(x) + g_2(x)) = 0.$$

Clearly, at points of non-differentiability it must also be decreasing.

step and G_ℓ has two steps in $[\tilde{v}_j, 1]$, for $j, \ell \in \{1, 2\}$, we adopt the following convention for β_j^2, Δ_j^2 (which are not well defined since G_j has only one step).

We define:

$$z_j \triangleq \max\{x \in suppG_j \backslash \{\beta_j^1\}\}. \tag{32}$$

That is, z_j is the highest element in the support of G_j taking our β_j^1. We adopt the following convention for β_j^2, Δ_j^2:

$$\beta_j^2 = \begin{cases} 0 & \text{if } z_j \leq \beta_\ell^2; \\ \frac{\beta_\ell^2 + z_j}{2} & \text{if } z_j > \beta_\ell^2, \end{cases} \tag{33}$$

$$\Delta_j^2 = \sup_{x < \beta_j^1} G_j(x) - \sup_{x < \beta_j^2} G_j(x). \tag{34}$$

In other words, Δ_j^2 is the mass probability of G_j in $[\beta_j^2, \beta_j^1)$. The important thing about the conventions adopted is the following. Let $k, h \in \{1, 2\}$ be such that $\beta_k^2 \geq \beta_h^2$, then by construction $\Delta_k^2 > 0$ and $suppG_h \cap (\beta_k^2, \beta_k^1) = \emptyset$.

Sub-sub-case (i) Throughout sub-sub-case (i), we relabel agents without loss of generality so that $\beta_1^2 \geq \beta_2^2$. We show that, if $\Delta_2^1 \leq 1/2$ or $\beta_3^1 \leq \beta_1^2$, the information structure is not optimal. Consider the information structure in which $\tilde{G}_2 = G_2$, $\tilde{G}_3 = G_3$ and

$$\tilde{G}_1(x) = \begin{cases} G_1(x) & \text{if } x < \beta_1^2 \\ G_1(\beta_1^2) - \sup_{x < \beta_1^2} G_1(x) & \text{if } \beta_1^2 \leq x < \mathbb{E}[v_1 \mid v_1 \in [\beta_1^2, 1]] \\ 1 & \text{otherwise} \end{cases}$$

In other words, the signal remains the same if the expected utility is below $\mathbb{E}[v_1 \mid s_1] < \beta_1^2$ and otherwise, all signals are pooled into one signal $\bar{s}$.

We have that:

$$R - \tilde{R} = \int_{\beta_1^2}^{\beta_1^1} (G_1(x) - \tilde{G}_1(x))(-G_2(x) - G_3(x) + 2G_2(x)G_3(x))dx. \tag{35}$$

To see why we get this expression, note that $\tilde{G}_2 = G_2$, $\tilde{G}_3 = G_3$ and $G_1(x) = \tilde{G}_1(x)$ for all $x \notin [\beta_1^2, \beta_1^1)$ so the terms in the integral outside this interval cancel out, so we just need to analyze the integral in the interval as it appears in (35).

Note that G_2 is constant in (β_1^2, β_1^1). Regarding G_3 we need to consider two cases. If $\beta_3^1 \leq \beta_2^1$, then G_3 is constant in (β_1^2, β_1^1); if $\beta_3^1 > \beta_2^1$ and $\Delta_2^1 \leq 1/2$, then $G_2(x) \geq 1/2$ in the interval (β_1^2, β_1^1), so

$(-G_2(x)-G_3(x)+2G_2(x)G_3(x))$ is non-decreasing in (β_1^2,β_1^1). Hence, $(-G_2(x)-G_3(x)+2G_2(x)G_3(x))$ is non-decreasing regardless of whether $\beta_3^1 \le \beta_2^1$ or $\beta_3^1 > \beta_2^1$. We also have that $\int_y^{\beta_1^1} G_1(x)dx \le \int_y^{\beta_1^1} \tilde{G}_1(x)dx$ with equality if $y=\beta_1^2$. We thus conclude that $R \le \tilde{R}$. However, information structure $\tilde{G}_1, G_2, G_3$, is such that $\tilde{\beta}_1^1, \beta_3^1 < \beta_2^1$, so Lemma 2 implies this is not an optimal information structure. Hence, we conclude that $\Delta_2^1 > 1/2$ is a necessary condition for an information structure to be optimal.

Also, note that if $\beta_2^2 = \beta_1^2$, then we could use the same argument to prove that if $\Delta_1^1 \le 1/2$, the information structure is suboptimal. Hence, a necessary condition for optimality (in addition to $\Delta_2^1 > 1/2$) is that $\beta_2^2 < \beta_1^2$ or $\Delta_1^1 > 1/2$.

Sub-sub-case (ii) We now adopt the convention (32)-(34), but assume that $\Delta_2^1 > 1/2$ and $\beta_3^1 > \max\{\beta_1^2, \beta_2^2\}$ and show that the information structure is optimal only if $\Delta_1^1 > 1/2$. As argued above, if $\Delta_1^1 \le 1/2$ and the information structure is optimal, then $\beta_2^2 < \beta_1^2$. We assume that $\Delta_1^1 \le 1/2$ and show the information structure is not optimal.

We first note that, if G_1 has only one step in $[\tilde{v}_1, 1]$, then we necessarily have that $\Delta_1^1 \ge \Delta_2^1 > 1/2$. Hence, if $\Delta_1^1 \le 1/2$, then G_1 has two steps in $[\tilde{v}_1, 1]$

We next prove that, if $\beta_2^2 < \beta_1^2$ and $\Delta_1^1 \le 1/2$ then the information structure is suboptimal. Now consider the information structure in which $\tilde{G}_2 = G_2$, $\tilde{G}_3 = G_3$ and

$$\tilde{G}_1(x) = \begin{cases} G_1(x) & x < \beta_1^2 - \varepsilon \\ G_1(\beta_1^2) - \eta & \beta_1^2 - \varepsilon \le x \le \beta_1^1 \\ 1 & \text{otherwise,} \end{cases}$$

where η is such that:

$$(G_1(\beta_1^2) - \eta)(\beta_1^1 - (\beta_1^2 - \varepsilon)) = G_1(\beta_1^2)(\beta_1^1 - \beta_1^2),$$

and ε is small enough so that $\beta_1^2 - \varepsilon > \beta_1^1$ and $\int_y^1 \tilde{G}_1(x)dx \ge \int_y^1 F(x)dx$ for all $y \in [\beta_1^2 - \varepsilon, 1]$. We then have that:

$$R - \tilde{R} = \int_{\beta_1^2}^{\beta_1^1} (G_1(x) - \tilde{G}_1(x))(-G_2(x) - G_3(x) + 2G_2(x)G_3(x))dx. \tag{36}$$

Note that $\tilde{G}_2 = G_2$, $\tilde{G}_3 = G_3$ and $G_1(x) = \tilde{G}_1(x)$ for all $x \notin [\beta_1^2, \beta_1^1)$ so the terms in the integral outside this interval cancel out, so we just need to analyze the integral in the interval as it appears in (36). Also, note that G_2 is constant in (β_1^2, β_1^1) while G_3 is strictly increasing at one point in this interval. However, since $\Delta_2^1 > 1/2$, we must have that $G_2(x) < 1/2$ in the interval (β_1^2, β_1^1). We then have that $(-G_2(x) - G_3(x) + 2G_2(x)G_3(x))$ is strictly increasing at one point in (β_1^2, β_1^1). We also have that $\int_y^{\beta_1^1} G_1(x)dx \ge \int_y^{\beta_1^1} \tilde{G}_1(x)dx$ with equality if $y = \beta_1^2$. We thus conclude that $R < \tilde{R}$.

Sub-case B Finally, we consider the case in which G_j has only one step in $[\tilde{v}_j, 1]$ for both $j \in \{1,2\}$. We note that in this case:

$$\Delta_1^1 = \Delta_2^1 = 1 - G(\tilde{v}_1) = 1 - G(\tilde{v}_1).$$

Hence, we assume that $\Delta_1^1 = \Delta_1^2 \leq 1/2$ and reach a contradiction. Following Theorem 2 in Kleiner, Moldovanu, and Strack (2021), we know that for $j \in \{1,2\}$ there exists a second interval $[\tilde{v}_j', \tilde{v}_j]$ such that G_j in this interval either: (a) is equal to F (i.e., there is complete information), (b) has one or two atoms and is constants everywhere else. There case in which G_j has one or two atoms in $[\tilde{v}_j', \tilde{v}_j]$ for some $j \in \{1,2\}$ implies the information structure is suboptimal, which can be proven in a completely analogous case sub-case 1A. We thus imply that G_j is equal to F in the interval $[\tilde{v}_j', \tilde{v}_j]$. We thus conclude that there exists a ε such that $G_1(x) = G_2(x)$ for every $x \in [\tilde{v}_1 - \varepsilon, 1]$, and the distributions have an atom of size smaller than $1/2$ at β_1^1.

We consider the following information structure:

$$\tilde{G}_j(x) = \begin{cases} G_j(x) & x \leq \tilde{v}_1 - \varepsilon \\ F(\tilde{v}_1 - \varepsilon) & x \in [\tilde{v}_1 - \varepsilon, \tilde{v}_1] \\ F(x - \varepsilon) & x \in [\tilde{v}_1, \tilde{v}_1 + \varepsilon] \\ F(\tilde{v}_1) & x \in [\tilde{v}_1 + \varepsilon, \beta_1^1 - \eta] \\ 1 & \text{otherwise,} \end{cases}$$

where η is such that:

$$\int_{\tilde{v}_1 - \varepsilon}^1 G_j(x)dx = \int_{\tilde{v}_1 - \varepsilon}^1 \tilde{G}_j(x)dx.$$

We can write this condition also as follows:

$$\eta(F(\tilde{v}_1) - 1) = \varepsilon(F(\tilde{v}_1 - \varepsilon) - F(\tilde{v}_1)) \tag{37}$$

We now prove that information structure $\tilde{G}_1, \tilde{G}_2, G_3$ generates higher revenue than G_1, G_2, G_3. We then can write the difference between the revenues generated as follows:

$$\begin{aligned} \tilde{R} - R = &\int_{\tilde{v}_1 - \varepsilon}^{\tilde{v}_1 - \varepsilon} -(\tilde{G}^2(x) + 2\tilde{G}(x)G_3(x) - 2\tilde{G}^2 G_3(x)) + (G^2(x) + 2G(x)G_3(x) - 2G^2 G_3(x))dx \\ &+ \int_{\beta_1^1 - \eta}^{\beta_1^1} -(\tilde{G}^2(x) + 2\tilde{G}(x)G_3(x) - 2\tilde{G}^2 G_3(x)) + (G^2(x) + 2G(x)G_3(x) - 2G^2 G_3(x))dx, \end{aligned}$$

where we used that $\tilde{G}(x) \triangleq \tilde{G}_1(x) = \tilde{G}_2(x)$ and $G(x) \triangleq G_1(x) = G_2(x)$ in the range of the intervals. We now use the expressions for $\tilde{G}$ and the fact that $G_3(\beta_1^1 - \eta) = 1$ to write the difference between revenues as follows:

$$
\begin{aligned}
\tilde{R} - R = \int_0^{\varepsilon} \Bigg(& \big(F^2(\tilde{v}_1 - \varepsilon + x) + 2G_3(\tilde{v}_1 + x)F(\tilde{v}_1 - \varepsilon + x) - 2F^2(\tilde{v}_1 - \varepsilon + x)G_3(\tilde{v}_1 + x)\big) \\
& - \big(F^2(\tilde{v}_1 - \varepsilon + x) + 2G_3(\tilde{v}_1 - \varepsilon + x)F(\tilde{v}_1 - \varepsilon + x) - 2F^2(\tilde{v}_1 - \varepsilon + x)G_3(\tilde{v}_1 - \varepsilon + x)\big) \Bigg) dx \\
& + \varepsilon \Bigg(\big(F^2(\tilde{v}_1) + 2\int_{\tilde{v}_1}^{\tilde{v}_1 + \varepsilon} G_3(x)dx(F(\tilde{v}_1) - F^2(\tilde{v}_1))\big) \\
& - \big(F^2(\tilde{v}_1 - \varepsilon) + 2\int_{\tilde{v}_1 - \varepsilon}^{\tilde{v}_1} G_3(x)dx(F(\tilde{v}_1 - \varepsilon) - F^2(\tilde{v}_1 - \varepsilon))\big) \Bigg) \\
& + \eta(F(\tilde{v}_1) - 1)
\end{aligned}
$$

Since G_3 is non-decreasing, we have that:

$$
\begin{aligned}
\tilde{R} - R \geq & \varepsilon \Bigg(\big(F^2(\tilde{v}_1) + 2G_3(\tilde{v}_1)F(\tilde{v}_1) - 2F^2(\tilde{v}_1)G_3(\tilde{v}_1)\big) \\
& - \big(F^2(\tilde{v}_1 - \varepsilon) + 2G_3(\tilde{v}_1)F(\tilde{v}_1 - \varepsilon) - 2F^2(\tilde{v}_1 - \varepsilon)G_3(\tilde{v}_1)\big) \Bigg) \\
& + \eta(F(\tilde{v}_1) - 1)
\end{aligned}
$$

Letting I denote the expression to the right of the inequality and taking derivatives of this expression with respect $G(\tilde{v}_1)$ we get:

$$
\frac{dI}{dG_3(\tilde{v}_1)} = 2((F(\tilde{v}_1) - F^2(\tilde{v}_1)) - F(\tilde{v}_1 - \varepsilon) - F^2(\tilde{v}_1 - \varepsilon)) \leq 0,
$$

where the inequality follows from the fact that $1/2 \leq F(\tilde{v}_1 - \varepsilon) < F(\tilde{v}_1)$, for ε small enough. Hence,

$$
\begin{aligned}
\tilde{R} - R \geq & \varepsilon \Bigg(\big(F^2(\tilde{v}_1) + 2F(\tilde{v}_1) - 2F^2(\tilde{v}_1)\big) && (38) \\
& - \big(F^2(\tilde{v}_1 - \varepsilon) + 2F(\tilde{v}_1 - \varepsilon) - 2F^2(\tilde{v}_1 - \varepsilon)\big) \Bigg) + \eta(F(\tilde{v}_1) - 1) && (39)
\end{aligned}
$$

Using (37) to replace the last term, we get that:

$$\begin{aligned}\tilde{R} - R \geq& \varepsilon\Bigg(\big(F^2(\tilde{v}_1) + 2F(\tilde{v}_1) - 2F^2(\tilde{v}_1)\big) \\ & - \big(F^2(\tilde{v}_1 - \varepsilon) + 2F(\tilde{v}_1 - \varepsilon) - 2F^2(\tilde{v}_1 - \varepsilon)\big)\Bigg) + \varepsilon(F(\tilde{v}_1 - \varepsilon) - F(\tilde{v}_1)) \\ =& 2\varepsilon\Bigg(\big(F(\tilde{v}_1) - F^2(\tilde{v}_1)\big) - \big(F(\tilde{v}_1 - \varepsilon) - F^2(\tilde{v}_1 - \varepsilon)\big)\Bigg) > 0,\end{aligned}$$

where we once again use that $1/2 \leq F(\tilde{v}_1 - \varepsilon) < F(\tilde{v}_1)$, for ε small enough. We thus conclude that $\tilde{R} > R$, and hence, G_1, G_2, G_3 is not an optimal information structure.

Case 2 We now assume that $\beta_1^1 = \beta_2^1 = \beta_3^1$ and prove that there exists $i, j \in \{1, 2, 3\}$ such that $\Delta_i^1, \Delta_j^1 \geq 1/2$. For each j let z_j be the maximum of the points in the support of G_j except for β_j^1 :

$$z_j = \max\{x \in suppG_j \mid x < \beta_j^1\}.$$

Suppose that there exists $z_j > z_k, z_\ell$ or $z_j \geq z_k, z_\ell$ and $z_j = \beta_j^2$ and $\Delta_j^2 > 0$. Then we can find an information structure that generates the same revenue and in which $\tilde{b}_j^1 < \beta_i^1, \beta_\ell^1$. Without loss of generality we assume that $j = 3$. Consider the information structure in which $\tilde{G}_1 = G_1$, $\tilde{G}_2 = G_2$ and

$$\tilde{G}_3(x) = \begin{cases} G_3(x) & \text{if } x \leq \beta_3^2 \\ G_2(\beta_3^2) - \Delta_3^2 & \text{if } \beta_3^2 \leq x < \frac{\Delta_3^2\beta_3^2 + \Delta_3^1\beta_3^1}{\Delta_3^2 + \Delta_3^1} \\ 1 & \text{otherwise} \end{cases}$$

In other words, the signal remains the same if the expected utility is below $\mathbb{E}[v_3 \mid s_3] < \beta_3^2$ and otherwise, all signals are pooled into one signal $\bar{s}$. We then have that:

$$R - \tilde{R} = \int_{\beta_3^2}^{\beta_3^1} (G_3(x) - \tilde{G}_3(x))(-G_1(x) - G_2(x) + 2G_1(x)G_2(x))dx. \tag{40}$$

Note that $\tilde{G}_1 = G_1$, $\tilde{G}_2 = G_2$ and $G_3(x) = \tilde{G}_3(x)$ for all $x \notin [\beta_3^2, \beta_3^1)$ so the terms in the integral outside this interval cancel out, so we just need to analyze the integral in the interval as it appears in (40). Also, note that G_1 and G_2 are constant in (β_1^2, β_1^1), so $(-G_1(x) - G_2(x) + 2G_1(x)G_2(x))$ is constant in (β_3^2, β_3^1). We also have that $\int_y^{\beta_3^1} G_3(x)dx \leq \int_y^{\beta_3^1} \tilde{G}_3(x)dx$ with equality if $y = \beta_3^2$. We thus conclude that $R = \tilde{R}$. However, information structure $G_1, G_2, \tilde{G}_3$, is such that $\beta_3^1 < \tilde{b}_1^1, \beta_2^1$, and in Case 1 we proved that this is optimal only if $\Delta_1^1 = \Delta_2^1 > 1/2$. ■

8.3.2 Local Optimality of the Symmetric Information Structures

Finally, we verify that, if we fix the information structure of $N-1$ bidders to be the optimal symmetric one and optimize over the information structure of the remaining bidder, we get the optimal symmetric information structure. In other words, it is not possible to generate a higher expected revenue by changing the information structure of only one bidder. Hence, the optimal symmetric information structure is a local optimum.

Proposition 10 (Local Optimality)
If the distribution of expected valuations of agents $\{1, ..., N-1\}$ is fixed to be the optimal symmetric information structure, then the information structure for bidder N that maximizes revenue is the optimal symmetric information structure.

Proof. Let $\hat{G}$ be the optimal symmetric information structure (characterized in Proposition 1). If the distribution of expected valuations of agent N is G, the probability that the second-order statistic is less than x is given by:

$$\mathbb{P}\{v_{(2)} \leq x\} = \hat{G}^{N-1}(x) + (N-1)\hat{G}^{N-1}(x)(1-\hat{G}(x))G(x).$$

We then have that the expected revenue is given by:

$$R = 1 - \int \hat{G}^{N-1}(x) + (N-1)\hat{G}^{N-1}(x)(1-\hat{G}(x))G(x)dx = -\int (N-1)\hat{G}^{N-1}(x)(1-\hat{G}(x))G(x)dx + \text{constants}.$$

where the term "constants" refers to terms that do not depend on G. Integrating by parts,

$$R = \int u(x)dG(x) + constants,$$

where

$$u(x) = \int_0^x (N-1)\hat{G}^{N-1}(y)(1-\hat{G}(y))dy.$$

We then have that the optimal information structure for agent N is given by:

$$G^* \in \arg\max_{F \prec G} \int u(x)dG(x). \tag{41}$$

This is a Bayesian persuasion problem as studied by Dworczak and Martini (2019) . More precisely, the maximization problem is a Bayesian persuasion problem where there is a continuum of states (in our model, a continuum of valuations), the sender's utility (in our model, the seller's revenue) only depends on the expected state induced by the signal (in our model, the distribution of expected valuation).

To begin, it is convenient to give the verification result found in Dworczak and Martini (2019):

Proposition 11 (Verification Theorem by Dworczak and Martini (2019))
If there exist a cumulative distribution function G and a convex function $\psi : [0,1] \to \mathbb{R}$, with $\psi(x) \geq u(x)$ for all $x \in [0,1]$, that satisfy

$$supp(G) \subset \{x \in [0,1] : u(x) = \psi(x)\} \tag{42}$$

$$\int_0^1 \psi(x) dG(x) = \int_0^1 \psi(x) dF(x) \tag{43}$$

$$F \text{ is a mean-preserving spread of } G, \tag{44}$$

then G is a solution to problem (41).

We use this verification theorem top show that $\hat{G}$ is a solution to (41).

We first recall that we can write $\hat{G}$ (the optimal symmetric information structure) as follows:

$$\hat{G} = \begin{cases} F(x) & x_2 \leq x \\ F(x_2) & x_2 \leq x \leq x_1 \\ 1 & \text{otherwise} \end{cases},$$

where

$$x_2 \triangleq F^{-1}(q) \quad \text{and} \quad x_1 \triangleq \frac{\int_{x_2}^1 x dF(x)}{1 - F(x_2)},$$

with q begin as in Proposition 1 . We constructing function ψ as follows:

$$\psi(x) = \begin{cases} u(x) & x \leq x_2 \\ u(x_2) + u'(x_2)(x - x_2) & x \geq x_2 \end{cases}$$

We make two observations. First, $\psi(x) = u(x)$ for all $x \leq x_1$ (note that $\hat{G}(x)$ is constant in (x_2, x_1) so $u(x)$ is affine in this segment). Second, $\psi(x)$ is convex. To verify the convexity, note that $\psi(x)$ is convex if and only if $u(x)$ is convex in $[0, x_2]$. However, taking the second derivative of $u(x)$ it is easy to verify that $u(x)$ is convex if and only if $x \leq (N-2)/(N-1)$ and by construction of the optimal quantile $x_2 = F(q) \leq (N-2)/(N-1)$ (see the proof of Proposition 1).

To verify that $\hat{G}$ is a solution to (41) we check that (42)-(44) are satisfied. First, $\hat{G}$ satisfies (44) because by construction the optimal symmetric information structure satisfies this condition. Second, note that $supp(\hat{G}) = [0, x_2] \cup \{x_1\}$ and as previously explained $u(x) = \psi(x)$ in this set, so (42) is also satisfied.

Finally, we have that:

$$\int_0^1 \psi(x)dF(x) = \int_0^{x_2} \psi(x)dF(x) + (1 - F(x_2))\psi(\frac{\int_{x_2}^1 \psi(x)dF(x)}{1 - F(x_2)}) \tag{45}$$

$$= \int_0^{x_2} \psi(x)dF(x) + (1 - F(x_2))\psi(x_1) = \int_0^1 \psi(x)d\hat{G}(x). \tag{46}$$

Hence, (43) is also satisfied. It follows that $\hat{G}$ is a solution to (41). ■