CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

Third-Degree Price Discrimination in the Age of Big Data

George Charlson
University of Cambridge

Abstract

A platform holds information on the demographics of its users and wants maximise total surplus. The data generates a probability over which of two products a buyer prefers, with different data segmentations being more or less informative. The platform reveals segmentations of the data to two firms, one popular and one niche, preferring to reveal no information than completely revealing the consumer's type for certain. The platform can improve profits by revealing to both firms a segmentation where the niche firm is relatively popular, but still less popular than the other firm, potentially doing even better by revealing information asymmetrically. The platform has an incentive to provide more granular data in markets in which the niche firm is particularly unpopular or in which broad demographic categories are not particularly revelatory of type, suggesting that the profit associated with big data techniques differs depending on market characteristics.

Reference Details

2159	Cambridge Working Papers in Economics
2021/04	Janeway Institute Working Paper Series
Published	16 August 2021
Key Words	Strategic interaction, network games, interventions, industrial organisation, platforms, hypergraphs
JEL Codes	D40, L10, L40
Websites	www.econ.cam.ac.uk/cwpe www.inet.econ.cam.ac.uk/working-papers

Third-Degree Price Discrimination in the Age of Big Data

George Charlson*

August 16, 2021

Abstract

A platform holds information on the demographics of its users and wants maximise total surplus. The data generates a probability over which of two products a buyer prefers, with different data segmentations being more or less informative. The platform reveals segmentations of the data to two firms, one popular and one niche, preferring to reveal no information than completely revealing the consumer's type for certain. The platform can improve profits by revealing to both firms a segmentation where the niche firm is relatively popular, but still less popular than the other firm, potentially doing even better by revealing information asymmetrically. The platform has an incentive to provide more granular data in markets in which the niche firm is particularly unpopular or in which broad demographic categories are not particularly revelatory of type, suggesting that the profit associated with big data techniques differs depending on market characteristics.

*Cambridge INET, Austin Robinson Building, Sidgwick Ave, Cambridge CB3 9DD, gc556@cam.ac.uk

KEYWORDS: Strategic interaction, network games, interventions, industrial organisation, platforms, hypergraphs. JEL classification: D4, L1, L4.

1 Introduction

Large online platforms hold increasingly detailed information about their users, which most commonly involves the collection of demographic data, such as age, location, gender and race, and behavioural data, such as browser history, device type and browser fingerprinting. The focus of the discussion on this form of so-called "big data" has been on potential privacy concerns (see, e.g., Aquisti, 2014; Tao et al 2019; Cecere et al 2017 and Belleflamme and Vergote, 2016) and third-degree price discrimination in the case of monopolies (Shiller, 2014; Townley et al, 2017; Esteves and Cerqueira, 2017 and Esteves and Resende, 2019), with the latter literature focusing on the fact that such data is informative of consumer preferences.

However, a well-known result in the price discrimination literature (Thisse and Vives, 1988, Bester and Petrakis, 1996, Chen, 1997, Fudenberg and Tirole, 2000, and Armstrong, 2005) is that firms having information that allows them to price discriminate in a horizontally differentiated market tends to reduce profits. As such, it is less clear how platforms that host firms can use demographic and behavioural data to increase aggregate profits directly, and why such data is (at least partially) shared with firms on platforms like Airbnb and Amazon Marketplace.

We examine the case where there are two firms, one popular and one niche, on a platform selling to a unit mass of buyers. Buyers are of two types, with a consumer of type i preferring firm i. There are more consumers who prefer the popular firm to the niche firm. The platform potentially has access to information on a range of buyer characteristics, which are informative of the preferences of consumers in the sense that they generate a known posterior distribution of types that differ from the prior type probability. The platform owner chooses which of these characteristics to reveal to the firms, generating a "segmentation" of the dataset. We allow the platform to reveal different segmentations of the data to each firm, with the aim of maximising aggregate profit. We show an example of a case where the firms are each shown different segmentations of the same dataset in Figure 1 below.[1]

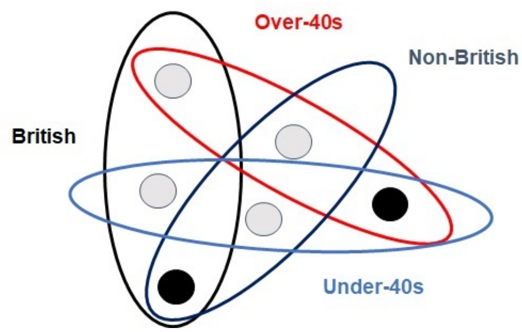

Figure 1: A representation of the case where one firm is shown information on buyer age, while the other is shown information on location. The black nodes represent firms and grey nodes represent buyers.

The platform owner faces a trade-off: segmentations that accurately reveal firm type increase the ability of firms to target firms of their own type, but reduce prices in segments where there are relatively few buyers of that type. Consistent with the price-discrimination literature cited above, fully revealing each buyers' type reduces profits relative to providing no information at all. However, profits can be increased by displaying segments that are partially informative: specifically, if there exists a segmentation

[1] As discussed in more detail below, throughout we will consider the platform owner's choice of segmentations as choosing a hypergraph, with each edge containing a segment of buyers and a single firm. To simplify the diagrammatic representation of the platform owner's problem further, we will show edges containing a small amount of buyer nodes, which, given the assumption that there is a unit mass of buyers, can be thought of as representing a number of buyers rather than a single buyer.

in which there are some segments in which the niche firm is more popular than they are on average, but still less popular than the other firm, then showing both firms such a segmentation is profit increasing. In this case, the price rise in segments where the niche firm is relatively more popular outweigh the fall in prices in segments where they are less so.

Furthermore, the platform is potentially able to increase profits further by showing the two firms different segmentations. Doing so can induce the niche firm to set a price that is only attractive to their own types in segments where the niche firm is relatively popular, which in turn increases the price both firms set in segments where the niche firm is less popular. Overall, this potentially causes a rise in prices for all consumers, increasing profits.

We also provide an account of the type of markets where big data is more likely to be profit increasing for the platform. In markets where broad demographic data is not a strong signal of preference, if the platform owner provides both firms with the same segmentation, then she has an incentive to provide both firms with more informative signals than in markets where broad signals are relatively informative of consumer preference.

Furthermore, the more niche the less popular firm is, the more informative the segmentation that the platform owner optimally provides: when a firm is particularly unpopular, it would require a very informative segmentation to induce it to set a higher price than the popular firm, and as such the platform can provide a more informative signal without inducing too much competition. This provides a potential explanation as to why niche products are becoming an increasingly large proportion of consumers' total consumption bundle in the last decade or so (Neiman and Vavra, 2020).

Our approach introduces two methodological innovations. First, we separate demographic information from buyer preference by having the platform choose segmentations according to the former in order to be informative about the latter. In contrast, approaches to similar problems (see Elliott, Koh and Galeotti, 2021 and Bounie, Dubus and Waelbroeck 2021) involve the platform directly segmenting up buyers according to their preferences directly. Our approach more closely captures the trade-offs involved in third degree price discrimination, where signals of preference are limited to the structure of the data held by the platform, and also make comparisons between markets with different levels of polarisation and distibutions of buyer preference.

Secondly, we show that an informative way of representing the platform owner's problem is that they are choosing between a set of hypergraphs that are consistent with some underlying data set. The hypergraphs are bipartite, linking a firm with a subset of buyers that share at least one characteristic, with firms choosing a different price for each edge of the hypergraph.[2],[3]

Along with the aforementioned work on price discrimination, our analysis fits into a wider literature on information design, contributions to which include Kamenica and Gentzkow (2011) and Bergemann and Morris (2016). A strand of such literature has explicitly examined how information design might be used by consumers (Ali et al, 2020), firms (Novshek and Sonnenschein, 1982; Vives, 1988, Raith, 1996; Johnson and Myatt, 2006) and platforms (Roesler and Szentes, 2017; Charlson 2020 and Armstrong and Zhou, 2020).

[2]A hypergraph is a generalisation of a graph in which the edges can be any non-empty subset of the nodes, as opposed to being only a pair of nodes.

[3]Economic papers that utilise hypergraphs in one form or another include those that model communication structures (Myerson, 1980; van den Nouweland et al., 1992; Slikker et al., 2000), those that model attack and defence networks (Dziubinski and Goyal, 2017) and network formation models (Chen, Elliott and Koh, 2020).

2 Model

Suppose there are two firms, f_1 and f_2, and a unit mass of buyers who interact on a platform. Firms offer a single, horizontally differentiated good. The platform is controlled by a platform owner, who provides information to the firms in order to maximise joint firm profits.

Suppose that a buyer b_i is associated with a set of base characteristics (e.g. age, race, gender, location). Let Ω be the set of characteristics known to the platform owner and a descriptor $D_i \subset \Omega$, refer to a collection of characteristics (such as "black, woman"), with $\boldsymbol{\mathcal{D}}$ denoting the set of all descriptors.

A segmentation, $\mathcal{S}_i$ of the set Ω is the set product of B and a descriptor D_i with cardinality y. A feasible segmentation can be denoted $\mathcal{S} = \{s_1, ..., s_y\}$ where: i) $s_i = \{b \in B | D_i \in \boldsymbol{\mathcal{D}}\}$; ii) $\{s_i \cap s_j\} = \emptyset$ for all $s_i, s_j \in \mathcal{S}$ and; iii) $\{s_1 \cup s_2 \cup ....s_y\} = B$. A segmentation partitions all buyers into a collection of segments of buyers that collectively contains every buyer and such that each segment contains a disjoint set of buyers that share a characteristic or set of characteristics, D.

Each buyer demands a single unit of a good inelastically, and there are no outside options. Let type space, $\boldsymbol{\upsilon} = \{1, 2\}$, be such that there are two types of buyer. Conditional on being of type i, a buyer, b_i, has a random valuation of $y_i \sim U[\frac{1}{2}, 1]$ for firm f_i's product and a valuation of $1 - y_i$ for f_j's product, where y_i is iid across buyers.[4] We define $\mathcal{V}_i$ as the set of type i buyers and let $\rho_i = \frac{|\mathcal{V}_i|}{n}$, the prior probability that a buyer is of type i. We assume throughout that f_1 is the "popular firm" with $\rho_1 > \frac{1}{2}$ that the other firm, f_2, is "niche".[5]

[4]This abstraction allows us to easily characterise the distribution of preferences of any arbitrary subset of buyers. The same principles of segmentation using data outlined here would still hold (and be more potent) if the platform owner had more precise information regarding the strength of buyer preference as well as its direction.

[5]The case where both firms are equally popular reduces to a Hotelling setting. In that case,

A segmentation, $\mathcal{S}_i$, then, along with the prior probability ρ_1 and the type space $\boldsymbol{\upsilon}$ generates a probability space $(\mathcal{S}_i, \boldsymbol{\upsilon}, \rho_1)$. The platform presents firm f_i with a segmentation, $\mathcal{S}_i$, of the data, for $i = 1, 2$, and thus its strategy is such that it chooses a set of two segmentations, $\{\mathcal{S}_1, \mathcal{S}_2\}$. A segmentation induces a posterior conditional probability distribution on $(\mathcal{S}_i, \boldsymbol{\upsilon}, \rho_1)$, $\alpha(\boldsymbol{\upsilon}|s_i)$,where $\alpha(\upsilon = 1|s_i)$ is the conditional probability that a buyer from the segment s_i is of type 1. We define a market, $\mathcal{M} = \boldsymbol{\mathcal{D}} \times \boldsymbol{\upsilon}$, as a distribution of buyers over type and description space.

The set of segmentations chosen and the subsequent posterior joint probability distributions generated is assumed to be common knowledge: f_i knows the distribution of types in across the segments within $\mathcal{S}_i$ and $\mathcal{S}_j$ but they do not know the precise valuation of their product by any one buyer, regardless of the segmentation chosen by the platform. Conditional on a set of segmentations $\{\mathcal{S}_1, \mathcal{S}_2\}$, let $\pi_{1k}(p_{1k}, \boldsymbol{p}_2; \mathcal{S}_1, \mathcal{S}_2)$ be f_1's expected profit from a segment k as a function of p_{1k} and the vector of 2's prices, $\boldsymbol{p}_2$. The firm f_1 receives demand, $x_{1i}(p_{1i}, \boldsymbol{p_2})$ from the segment s_i as follows:

$$x_{1i}(p_{1i}, \boldsymbol{p_2}) = |s_i| \sum_j [\alpha(\upsilon = 1|s_i, s_j)(1 - \chi(p_{1i} - p_{2j}))p_{ik} + \alpha(\upsilon = 2|s_i, s_j)\chi(p_{2j} - p_{1i})p_{1i}]$$

where $\chi(p_{1i} - p_{2j}) = p_{1i} - p_{2j}$ iff $p_{1i} - p_{2j} > 0$, and 0 otherwise. Hence the firm's maximisation problem for the segment s_i can be stated: $\max_{p_{1i}}\{p_{1i}x_{1i}(p_{1i}, \boldsymbol{p_1})\} = \max_{p_{1i}}\{\pi_{1i}(p_{1i}, \boldsymbol{p_2})\}$.The equilibrium of the pricing game is a a set of price vectors, $(\boldsymbol{p_1}, \ \boldsymbol{p_2})$, one for each firm, such that the price, p_{ik}, for a segment, s_k, is a best-response to the equilibrium price vector, $\boldsymbol{p}_j$, where $j \neq i$. We assume throughout that

the platform owner would require information regarding the strength of buyer preference in order to increase profits using segmentation, but the principles of doing so would be very similar to those we outline here.

the platform owner maximises joint expected profits, $\pi_P(\boldsymbol{p_1}, \boldsymbol{p_2})$, choosing a set of segmentations, $\mathcal{S}_1, \mathcal{S}_2$ to do so.

Let $p_{ik}(\boldsymbol{p}_j)$ denote f_i's reaction function, which is derive by finding the first order condition of the above maximisation problem. Then an equilibrium of the pricing game is such that $p^*_{ik}(\boldsymbol{p}^*_j)$ is a best reply to $\boldsymbol{p}^*_j$ for all i, j, k. Given the functional form of the demand curves here, here is a unique pure-strategy Nash equilibrium, $(\boldsymbol{p^*_1}, \boldsymbol{p^*_2})$ for any pair of segmentations, $(\mathcal{S}_1, \mathcal{S}_2)$.

3 The structure of information and hypergraphs

Hypergraphs

For a given market, $\mathcal{M}$, a pair of segmentations $(\mathcal{S}_{1,}\mathcal{S}_2)$ generates a bipartite hypergraph, $H = (B; f_1, f_2; \boldsymbol{E})$ where each $E_k \in \boldsymbol{E}$ is such that $E_k = \{f_i \cup s_k\}$ for one firm, f_i. We define a hypergraph as feasible if it can be generated by a feasible set of segmentations as defined above, and let Λ be the set of all feasible hypergraphs. Finally, let $\alpha(\boldsymbol{v}|s_k) = \alpha(\boldsymbol{v}|E_k)$ denote the conditional probability distribution generated by an edge, $E_k = \{f_i \cup s_k\}$.

As an example, suppose the segmentation is such that f_1 is shown all British people and non-British people separately, while f_2 is shown under-40s and over-40s separately. Such a situation can be represented in a hypergraph as depicted in Figure 1 above.

A firm, f_i's pricing problem can then be represented as f_i choosing a price p_{ik} for an edge, E_k, such that $f_i \in E_k$. The platform owner's problem can then be restated as choosing a hypergraph $H \in \Lambda$ which maximises joint profits. Throughout, we will consider the platform owner's problem in this way. As Figure 1 shows, considering the platform's problem in this way leads to an intuitive representation of different segmentation outcomes.

The information structure

We use the hypergraph representation described above to define way of comparing different segmentations of the data. We first define the union of two hypergraphs, H_1 and H_2 such that $H_1 \cup H_2 = \{B; f_1, f_2; E_1 \cup E_2\}$ and $H_i = \{\hat{B}; f_1, f_2; \{\hat{E}_i\}_i\}$ as a subhypergraph of $H = \{B; f_1, f_2; \{E_i\}_i\}$ if $\hat{B} \subseteq B$ and $\hat{E}_i \subseteq \{E_i \cap \hat{B}_i\}$ for all $\hat{E}_i$: a subhypergraph of H is thus created from H by the deletion of vertices that are elements of H.

We then define H' as a "feasible refinement" of H if H' is feasible and H' can be generated by taking the union of some subhypergraphs of H, i.e. $H' = \{\cup_{H_i \in \Upsilon} H_i\}$, where Υ denotes a set of subhypergraphs of H. We show an example of a refinement of a hypergraph below:

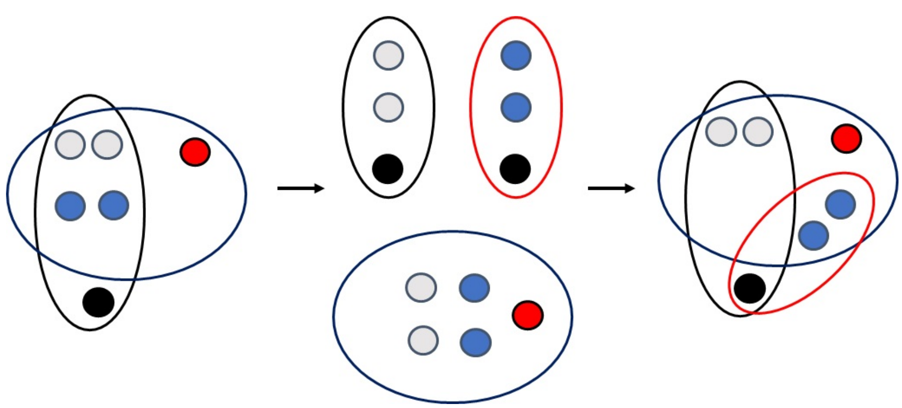

Figure 2: The original hypergraph H is composed of three subhypergraphs in the middle picture, the union of which is a refinement of H, H'.

The concept of a refinement is a natural way of ordering the information structures generated by demographic data. The hypergraph generated by showing both firms, for example, data on the gender and race of every buyer would be a refinement on the hypergraph generated by showing them data on just gender or just race. Every feasible hypergraph generated by the platform is a refinement on the no-information hypergraph, H_n, in which there are two edges: $E_1 = \{f_1 \cup B\}$ and $E_2 = \{f_2 \cup B\}$. Similarly, the complete-information hypergraph, H_c, with edges, $E = \{f_i \cup b_j\}$ for each firm-buyer pair f_i, b_j, is a refinement of any feasible hypergraph.

Let y_i denote the number of segments in the segmentation $\mathcal{S}_i$. Let $P(\mathcal{S}_i)$ denote the information structure generated by $\mathcal{S}_i$ such that $P(\mathcal{S}_i)$ is a $2 \times y_i$ right-stochastic matrix whose generic element is $\beta_{jk}(\mathcal{S}_i) = \Pr(b \in E_j | b \in \mathcal{V}_k, \mathcal{S}_i)$. Let R be a garbling (Markov) matrix whose rows are probability vectors. We state the following definition, following Blackwell (1953):

Definition. $\mathcal{S}_i^{'}$ is more informative than $\mathcal{S}_i$ iff $R^T P(\mathcal{S}_i^{'}) = P(\mathcal{S}_i)$.

Following this definition, we state that $H^{'}$ is more informative than H iff the segmentations that generate $H^{'}$,$\mathcal{S}_1^{'}$ and $\mathcal{S}_2^{'}$, are such that $\mathcal{S}_1^{'}$ and $\mathcal{S}_2^{'}$ are more informative than $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively, where the latter two segmentations generate H. We state the following result that reflects the informativeness of a refinement of a hypergraph, $H^{'}$, compared with H itself:

Proposition 1. *For a given market, if $H^{'}$ is a refinement of H, then $H^{'}$ is more informative than H.*

Proof. For $H^{'}$ to be a refinement of H, it must be that $D^{'}P(\mathcal{S}_i^{'}) = P(\mathcal{S}_i)$ for $i = 1, 2$, where $D^{'}$ is a matrix filled only with 1s and 0s and whose ith row and jth column components are such that $\sum_i d_{ij} = 1$. This follows as for any edge, E_k, of the original hypergraph, H, there is a set of edges, Φ_k, with generic element, $E_k^{'} \in H^{'}$, such that collectively the elements of Φ_k contain every element of E_k. Hence, $\sum_{k=1}^{|\Phi_k|} \beta_{jk}(\mathcal{S}_i^{'}) =$

$\beta_{jk}(\mathcal{S}_i)$. The matrix D then satisfies the criteria for it to be a garbling matrix, which immediately implies the Proposition. $\square$

A refinement of H by definition provides more granular demographic data than H. Intuitively, that refinement is at least as informative as H, because at the very least the firm can ascertain the distribution of types in an original segment of consumers in H from the subset of segments that it is split into in $H^{'}$. Our definition of informativeness is thus a useful way of comparing different segmentations shown to firms by the platform owner in the context being considered.

4 The costs and benefits of informativeness

We characterise how the selective provision of information on consumer type may increase profits from the benchmark of the no-information hypergraph. To do so, we first identify a necessary condition for a segmentation to improve upon the no-information hypergraph. For this to be the case, it must be that the platform is able to induce at least one of the firms to target more of their own types than in the no-information case.

Define $\mathcal{S}$ as informative iff it is more informative than $\mathcal{S}_n$. Then the following result holds:

Proposition 2. *If $H_n \notin \Lambda^*$, then $\exists \mathcal{S}$ which is informative.*

If $\mathcal{S}$ is not informative, then $\alpha(\boldsymbol{v}|E_k)$ for all $E_k \in H$. In this case, the equilibrium price vector associated with H, $(\boldsymbol{p_1^*}, \boldsymbol{p_2^*})$, is such that $f_i^{\prime}s$ price for any edge $E_k \in H$ $p_{ik} = p_n$,and hence $\pi_P(\boldsymbol{p_1^*}, \boldsymbol{p_2^*}) = \pi_P(p_{1n}, p_{2n})$, where p_{in} is the price f_i sets in response to H_n.

For the platform to be able to improve its profit from the no-information case, it must be that a buyer's type is not independent of a feasible descriptor, or, in other words, the conditional probability distribution $\alpha(\boldsymbol{v}|E_k)$ is not equal to the prior distribution of types, $\alpha(\boldsymbol{v})$. If not, it must be true that any segmentation produces a hypergraph that yields the same prices and profits as the no-information hypergraph. For third-degree price discrimination to be profitable (or indeed, possible), demographic information must be more informative of the type distribution of at least some buyers compared with providing no demographic information.

However, the condition in Proposition 1 is a necessary but not sufficient condition for third-degree price discrimination to be profitable. Providing a more precise signal to a firm f_i of each buyer's type increases the extent to which f_i competes for buyers of type j, which decreases prices for both firms. Hence, the platform owner faces a trade-off between using demographic data to reveal information and competition concerns.

To illustrate a case in which information revelation can be harmful to profits, we compare the no-information hypergraph to the complete information hypergraph, which is totally informative of each buyer's type, trivially satisfying the condition in Proposition 1. Letting $\succ$ represent the preferences of the platform owner, the following result holds:

Proposition 3. *Suppose $\rho_2 \neq 0$. Then, $H_n \succ H_c$.*

The result of Proposition 3 is consistent with the findings in Thisse and Vives (1988) and Bester and Petrakis (1996), whereby providing firm in a Hotelling setting with information on whether buyers prefer them or their competitor results in lower prices and profits. While information on buyer preference allows for price discrimination, the increased incentive for firms to compete hard for buyers of the other firm's type outweighs the profit increasing effect of differential pricing, reducing profits.

Profit increasing segmentations

Define a segmentation where $\alpha(v = 1|s_k) > \alpha(v = 2|s_k)$ for all $s_k \in \mathcal{S}$ as being "dominance-preserving" and let $H_{\mathcal{S}}$ refer to a "symmetrically informative" hypergraph where both firms are shown the same segmentation, $\mathcal{S}$. We can characterise a condition that allows comparison between a subset of the segmentations potentially available to the platform owner:

Theorem 1. *If $\mathcal{S}$ and $\mathcal{S}^{'}$ are both dominance-preserving and $\mathcal{S}$ is more informative than $\mathcal{S}^{'}$, then $H_{\mathcal{S}} \succ H_{\mathcal{S}^{'}}$.*

If a segmentation, $\mathcal{S}$, is more informative than another, $\mathcal{S}^{'}$, then the posterior distribution of types generated by $\mathcal{S}$ is a mean-preserving spread of the distribution generated by $\mathcal{S}^{'}$ (Blackwell, 1953). Conditional on f_i being more popular than f_j in every segment of a segmentation shown to both firms, the platform owner's profit function is convex in $\alpha(v = j|s_k)$ for every $s_k \in \mathcal{S}$. This implies that the platform can generate more profits by showing both firms the segmentation $\mathcal{S}$ than by showing both of them $\mathcal{S}^{'}$.

If $\mathcal{S}$ is an informative, dominance-preserving segmentation, then by definition it generates a posterior distribution $\alpha(\boldsymbol{v}|s)$ which is a mean-preserving spread of the type distribution $\alpha(\boldsymbol{v})$. As the hypergraph H_n is symmetrically informative, Theorem 1 immediately implies the following statement:

Corollary 1. *If there exists an informative dominance-preserving segmentation then there exists a H such that $H \succ H_n$.*

If there exists an informative dominance-preserving segmentation, then showing both firms that segmentation is preferred by the platform owner to providing them with no information at all. Note that Corollary 1 is a sufficient but not a necessary condition for it to be profitable for the platform owner to segment the market in this way. Segmentations that render the niche firm more popular in some segments can be profit increasing if they are sufficiently small so as not to reduce prices in other segments to too greater extent. Either way, the platform owner increasing its profits from the no-information case is possible if there exists a segmentation that is, in some sense, not too informative of consumer type. We provide an example that illustrates the result in Corollary 1 below.

Example 1

Assume that $\rho_1 = 0.8$ and buyers are either urban-dwellers or rural-dwellers according to the only data the platform owner holds. The marginal distribution of types is below:

	Type 1	**Type 2**	**Total**
Urban	0.3	0.2	0.5
Rural	0.5	0	0.5

Let s_u denote a segment containing every urban-dweller and s_r denote a segment containing every rural-dweller, such that $s_r \cup s_u = B$. Consider a 4-edge hypergraph, H, where $E_1 = \{f_1 \cup s_r\}, E_2 = \{f_1 \cup s_u\}$ with E_3 and E_4 defined analogously for f_2.

The equilibrium price profile generated by H is such that $p^*_{24} \approx \frac{7}{9} > 0.5 = p_{2n}$: firm 2 is induced to set a higher price than they do in H_n to these firms. This induces firm 1 to set a higher price to these consumers (in E_2) as well, setting a price of $\frac{8}{9}$ rather than $\frac{3}{4}$. On the other hand, $p_{11} = \frac{2}{3}$ and $p_{13} = \frac{1}{3}$. These prices imply that aggregate profit for the no-information hypergraph is 0.65, whereas it is more than 0.69 for H'. Hence, $H' \succ H_n$.

5 The optimal information structure

The preceding analysis indicates that full information relevation is always suboptimal for the platform, but under quite general conditions they are able to improve their payoff by revealing some information to the firms. The proof of Theorem 1 relied on the platform owner showing each firm the same segmentation. However, as we have seen, our framework allows for firms to be shown different segmentations to one another. In this section, we will explore more general information structures in order to analyse the platform owner's optimal segmentation choice.

A principle of least information revelation

We have shown that information revelation imposes a cost on the platform owner in the form of increased competition, while potentially increasing profits by inducing firms to set higher prices to groups of consumers who are relatively more likely to have a preference for their product. We identify cases in which further information revelation will not be beneficial to the platform owner in any circumstance.

Information revelation increases competition by inducing a firm to decrease their prices to a set of buyers who are less likely to be of their type. If this is not offset by price increases for other groups of consumers, overall profit must fall. It follows that providing more information to a firm that is setting a price that only attracts buyers of its own type, without providing more information to the other firm will decrease aggregate profits.

We formalise this intuition as follows. We define an edge, $E_k \in H$, as being targeted by a firm f_i if the probability that any buyer, $b \in E_k$, of type $j \neq i$ buys from f_i is zero in equilibrium for the set of information segmentations summarised by the hypergraph,

H. The following result holds:

Proposition 4. *Suppose that E_s is targeted by f_i and that for every $b \in E_s$, $b, f_j \in E_t$. If $H^{'}$ is a refinement of H created by taking the union of two or more subhypergraphs of E_s, with every other edge the same as in H, then $H \succeq H^{'}$.*

Proposition 4 establishes the principle that, conditional on buyers fitting a particular description are being targeted by a firm, and are observing a single price from the other firm, the platform has an incentive to give the first firm as little information on those buyers as possible. By providing more detailed information, the platform owner risks inducing the firm to set a price that attracts buyers of both types, which is unprofitable overall. By inducing a firm to set a price which only attracts buyers of its own type within a segment, the platform induces the firm to essentially ignore buyers of the other type. This increases prices throughout the market, and ensures more buyers purchase the product that yields them the highest gross surplus.

To give an example of the implications of this result, suppose that f_j is given information such that buyers are split between two edges based on whether they are women or men and that f_i is given information on buyer's gender and race. If, in equilibrium, f_i's price is such that no type j black women buy from them, then providing more information about those buyers (such as online behaviour or geographical data) would be suboptimal.

Asymmetric data revelation

Proposition 4 also implies that if f_i is targeting every buyer within a set of segments shown to the other firm (and therefore f_i's price is higher than any price set by f_j to buyers in these edges), then it is optimal (all else equal) for f_i to be shown the same segmentation as f_j. Doing so increases prices in edges where there is a more even distribution of types, more than it decreases prices in edges where f_i is more popular. ,

However, asymmetric segmentations can induce a firm that is less popular in a particular segment of buyers to set a price where only buyers of their own type buy from them. This outcome is possible because the firm, f_i, who is more popular in that segment can be induced to set a lower price than they otherwise would as they have to compete with f_j in another segment where there are even more type is.

The availability of asymmetric segmentations then present another trade-off for the platform owner. When firms are provided with symmetric segmentations, firms set higher prices in segments where types are more evenly distributed, which increases profits conditional on the segmentations not being too informative. With asymmetric segmentations, some of the benefits associated with more even type distributions are lost, because they generally involve one of the firms setting a single price in response to two or more prices set by the other firm. At the same time, such segmentations can induce the firm with more precise information to set a price such that only their own types buy from them, which in turn increases the price of the other firm, including in segments where type distribution is more unequal. This can increase profits for the platform owner, as Theorem 2 makes clear:

Theorem 2. *For any $\rho_1 \in (\frac{1}{2}, 1)$, there exists a dominance-preserving segmentation, $\mathcal{S}$, such that $(\mathcal{S}_n, \mathcal{S})$ generates a hypergraph, $H \succ H_{\mathcal{S}}$.*

Even if $\mathcal{S}$ and $\mathcal{S}'$ are dominance-preserving, it is still possible to induce f_2 to set a price seen by buyers in some edge $E \in H$ which only attracts type 2s. An asymmetrically informative hypergraph enables this as f_1 can be induced to set a lower price to some edge containing a relatively large number of type 2s than they would set in $H_{\mathcal{S}}$. This can induce f_2 to set a price higher than f_1 to these buyers, which can potentially increase profits.

To see how asymmetric information can lead to an increase aggregate profit, suppose f_1 is totally uninformed. Inducing f_2 to set a price in some edge, $E_k = s_k \cup f_2$, higher than they would in the equivalent edge in $H_{\mathcal{S}}$ (at least) reduces the losses associated with not revealing this segmentation by inducing f_1 to set a relatively high price to other segmentations. In fact, as f_1 is in the position of setting a lower price in $E_k^{'} = s_k \cup f_1$, but by definition there are more type 1s than type 2s in $E_k^{'}$, it is possible to induce f_1 to set a price which is greater than the price set by f_2 in $E_k \in H_{\mathcal{S}}$. This implies in turn that f_2 sets a higher price in $E_k \in H$ than f_1 sets in $E_k^{'} \in H_{\mathcal{S}}$. In which case, every buyer in H observes a pair of prices that are both higher than the prices that the buyer observes in $H_{\mathcal{S}}$. It follows that the platform's profit increases.

Theorem 2 then implies that the ability for the platform to provide firms with asymmetric data can be profit increasing relative to the case where they are restricted to symmetric data revelation: they are able to induce the niche firm to set a price that is higher than it would be for any symmetric cut of the data, which in turn can lead to an increase in profits overall. Further increases in profit can result in more complex cases such as when f_1 is induced to set a price higher than f_2's to a subset of the consumers that f_2 targets when f_1 is shown no information. We provide an example of this below.

Example 2

Suppose that the platform has data on the age of buyers. Assume that $\rho_1 = 0.8$, and that the buyer type marginal distributions can be represented by the following table:

	Type 1	**Type 2**	**Total**
Under-25	0.12	0.04	0.16
25-30	0	00.4	0.04
30+	0.68	0.12	0.8

First, we compare the hypergraph, $H_{\mathcal{S}}$ where both firms are shown a segmentation, $\mathcal{S}$, which places all thirty and under buyers into one segment (s_1) and all over thirty buyers into another (s_2), with H, where f_1 is uninformed and f_2 is shown the segmentation $\mathcal{S}$.

Let $E_1 = \{s_1 \cup f_1\}$, $E_2 = \{s_1 \cup f_2\}$, $E_3 = \{s_2 \cup f_1\}$, $E_4 = \{s_2 \cup f_2\}$ and $E_B = \{B \cup f_1\}$. Let $\boldsymbol{p_i^*}$ and $\boldsymbol{p_i^{'}}$ refer to f_i's price vector for $H_{\mathcal{S}}$ and H respectively. Then $p^{'}_{1B} \approx 0.79 > \frac{7}{9} = p^*_{21}$, $p^{'}_{21} \approx 0.89 > \frac{8}{9} = p^*_{12}$ and $p^{'}_{22} \approx 0.48 > 0.45 \approx p^*_{22}$. It follows that every buyer observes a set of prices in H strictly larger than the prices they observe in $H_{\mathcal{S}}$, and hence $H \succ H_{\mathcal{S}}$.

In fact, a further improvement can be made by showing a segmentation, $\mathcal{S}^{'}$ to f_1, with $\mathcal{S}^{'}$ splitting all under-25s into one segment and every other buyer into another. Let $H^{'}$ represent the hypergraph generated by the segmentations ($\mathcal{S}^{'}$,$\mathcal{S}$). $\mathcal{S}^{'}$ increases the price f_1 sets to the under 18s relative to the price it sets when it receives no information. While the price f_1 charges every other buyer decreases, the net effect is to increase profits, and so $H^{'} \succ H$. We depict this example in the figure below.

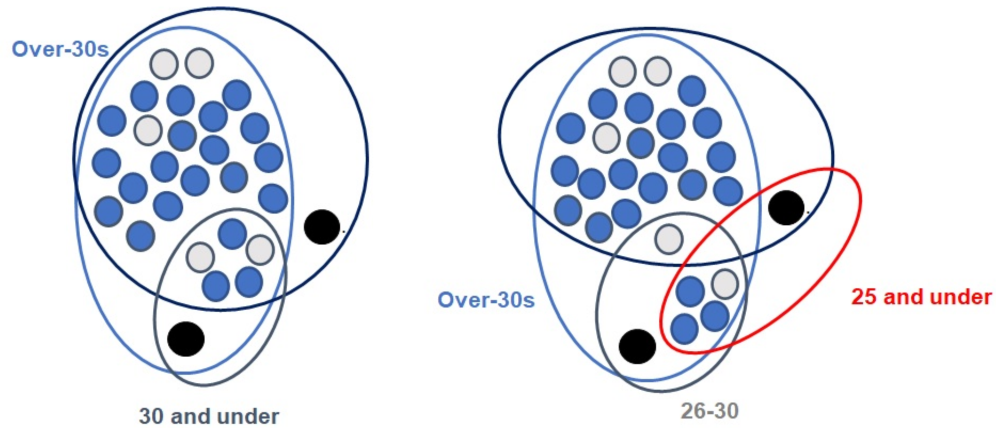

Figure 3: Blue and grey nodes represent type 1s and 2s respectively. Left hand panel: The platform can increase profits relative to symmetric segmentations by giving f_1 (on the right) no information, while providing f_2 with information on buyer age. Right hand panel: the platform can increase profits even more by providing $f_1$with a cut of the data which is informative, but different to the one shown to f_2.

6 Big data in different markets

Given the analysis in the previous sections, it is clear that the platform's incentive to reveal information varies depending on the distribution of buyer types across characteristics. We formalise this intuition by examining how market characteristics effect of the extent to which data is relevatory to that firm's type and the overall popularity of each firm's products.

Polarisation

Consider two markets, $\mathcal{M}_A$ and $\mathcal{M}_B$, each populated with two firms, but with potentially different distributions of buyers. We state the following definition:

Definition. Suppose that for two markets, $\mathcal{M}_A$ and $\mathcal{M}_B$: (1) $\rho_i^A = \rho_i^B$ for $i = 1, 2$

and; (2) for any set of feasible segmentations, $(\mathcal{S}_1, \mathcal{S}_2)$, if $\alpha(v = i|s_k, \mathcal{M}_A) \geq (\leq)\rho_i$, then $\alpha(v = i|s_k, \mathcal{M}_B) \geq (\leq)\rho_i$ and $\alpha(v = i|s_k, \mathcal{M}_A) \geq (\leq)\alpha(v = i|s_k, \mathcal{M}_B)$ for $i = 1, 2$. Then $\mathcal{M}_A$ is more polarised than $\mathcal{M}_B$.

When a market is more polarised, demographic data is more revealing of buyer type for any possible segmentation of the data set. For example, the demand for clothing is likely to be dependent on even relatively crude demographic data, like gender or race, whereas other markets, e.g. the market for white goods could be thought of as being rather less driven by such relatively broad categories. Note that in the latter type of markets, it would still generally be possible to accurately reveal consumer types to the firms; for example, by disclosing information on customer online behaviour a platform may be able to signal to firms which consumers are likely to value their products highly in a way they cannot using traditional demographic data. The following result holds:

Theorem 3. *Suppose that market $\mathcal{M}_A$ is more polarised than $\mathcal{M}_B$ and that $H_A \in \Lambda_A$ and $H_B \in \Lambda_B$. If H_A and H_B are symmetrically informative and dominance-preserving then H_A is not more informative than H_B under the type distribution in $\mathcal{M}_B$.*

When a market is less polarised than another, the platform owner has an incentive to provide firms with a segmentation that would be, in the more polarised market, more informative than the actual segmentation shown to firms in that market, if the segmentation is dominance-preserving in the original market. This follows because, as in Theorem 1, if $\mathcal{S}_A$ is a mean-preserving spread of $\mathcal{S}$ and is dominance-preserving then it generates higher aggregate profits than $\mathcal{S}$. This relationship necessarily holds in market $\mathcal{M}_B$, which implies the result - showing both firms a segmentation at least as informative as $\mathcal{S}_A$ is optimal, and hence showing both firms a segmentation which is too informative in $\mathcal{M}_A$ may be optimal in $\mathcal{M}_B$.

Theorem 3 suggests platforms have a greater incentive to invest in the collection of detailed behavioural and demographic data in markets that are less polarised when they are restricted to symmetric data revelation. In such markets, more general demographic insights are less likely to allow firms to accurately identify buyers who value their product highly, and thus are unable to set higher prices to profit from such buyers. Furthermore, the cost of revealing such data is reduced in less polarised markets, because the niche firm is more popular in segments in which they are relatively unpopular in the less polarised when compared with the more polarised. This softens the competition effect associated with releasing more granular data.

In the more general case where the platform provides firms with different segmentations, a less polarised market does not always result in both firms being provided with more information. For example, it might be that there is a segmentation that is too informative in the more polarised market which, in the less informative market, induces the niche firm to target their own firms when the popular firm is uninformed. This may be profit increasing relative to some symmetrically informative hypergraph. In this case, the niche firm is observing a more informative segmentation, but the popular firm is less informed. We illustrate this point in Example 3.

Nevertheless, Theorem 3 implies that as markets become less polarised, more detailed data on consumers will be useful to the platform owner, as such data is less likely to be "too informative" in such markets when compared with more polarised ones.

Example 3

Suppose that platform owner has information on gender (man or woman) and online behaviour (whether or not the buyer has searched for f_2's product online).We assume that $\rho_1 = 0.70$ and that there are a equal number of men and women, and not searched and searched, with each subcategory (e.g. men who have searched for the product online) making up 25% of the total proportion of buyers. We show the proportion of type 1s in each subcategory in two markets, $\mathcal{M}_A$ and $\mathcal{M}_B$, in the table below:

	$\boldsymbol{\mathcal{M}}_A$		$\boldsymbol{\mathcal{M}}_B$	
	Searched	**Not-searched**	**Searched**	**Not-searched**
Man	0.9	0.9	0.75	0.8
Woman	0.3	0.7	0.55	0.7

The distribution of firms in the two markets is also depicted in the left-hand panel of Figure 4.

By the definition of polarisation, $\mathcal{M}_A$ is more polarised than $\mathcal{M}_B$. Let H_G represent the hypergraph in which both firms are provided with data on the buyers' gender, H_D be the hypergraph where both firms are shown data on gender and on online behaviour and H be generated by showing f_1 no data and f_2 data on both gender and on online behaviour. In $\mathcal{M}_A$, the optimal hypergraph $H_G \in \Lambda^*_A$: both H_D and H involve too much revelation of type, and therefore do not yield as much profit as H_G.

In $\mathcal{M}_B$, $H_D \succ H_G$: in this case, both H_D, H_G are dominance-preserving and H_D yields more profit than H_G. This is an example of the result in Theorem 3. However, note that $H \succ H_D$: the platform owner prefers to provide less information to the popular firm in the less polarised market, as doing so allows them to show more information to f_2, inducing the latter to set a targeted price to women who have searched for f_2's product. In this case, f_2 is being optimally provided with more detailed data than before, though f_1 is being provided with less. We depict this in the right-hand panel of Figure 4.

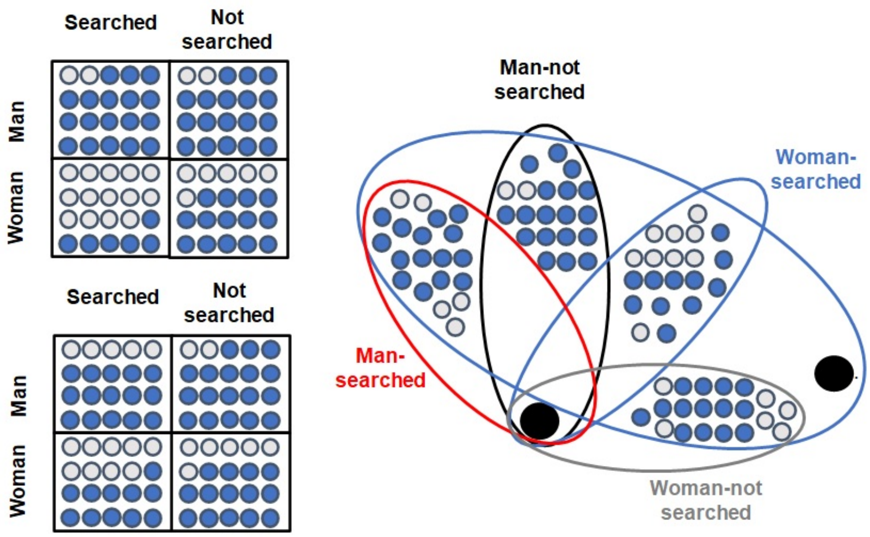

Figure 4: Left hand panel: the distribution of buyer types in $\mathcal{M}_A$ (above) and $\mathcal{M}_B$ (below). Right hand panel: the asymmetrically informative hypergraph H.

Niche firms

In this section, we consider how the relative popularity of firms and the distribution of that popularity affects the optimal revelation decisions of the platform owner. To do so, we define a measure of popularity which allows us to easily compare across markets:

Definition. f_i is uniformally less popular in market $\mathcal{M}_A$ than market $\mathcal{M}_B$ iff for any $\mathcal{S}_i$, $\alpha(v = i|s_k, \mathcal{M}_B) = \gamma\alpha(v = i|s_k, \mathcal{M}_A)$ for all $s_k \in \mathcal{S}_i$ and $\gamma \in (0, 1)$.

If f_i is uniformly less popular in market $\mathcal{M}_A$ than in market $\mathcal{M}_B$, then the proportion of type is in every feasible segmentation of the data is smaller in $\mathcal{M}_A$ than in $\mathcal{M}_B$. We state the following condition that links the popularity of the niche firm and the informativeness of the platform owner's optimal segmentation:

Proposition 5. *Suppose that f_2 is uniformly less popular in $\mathcal{M}_B$ than $\mathcal{M}_A$ and that $H_A \in \Lambda_A^*$ and $H_B \in \Lambda_B^*$. If H_A and H_B are symmetrically informative and dominance-preserving then H_A is not more informative than H_B under the type distribution in $\mathcal{M}_B$.*

If f_2 is uniformly less popular in one market than another, then each edge containing f_2 contains fewer type 2s by definition. The platform owner then has an incentive to reveal more granular information about buyer preferences in order to present the firm with a segment which contains a higher proportion of type 2s, with the aim of increasing f_2's ability to target buyers of their own type without reducing prices for other buyers to too greater a degree.

Proposition 5 implies that firms selling more niche products in certain markets are the first-order beneficiaries of the rise in big data techniques that allow the collection of more granular consumer data. This finding tallies with a body of empirical research that highlights the rise of niche products and product variety over the last one or two decades (see, Neiman and Vavra, 2020) One explanation for such a rise is that it has coincided with the popularity of both big data techniques and online markets more generally, which have allowed the collection of more data on consumers and better analysis of that data, which would be of particular benefit to niche firms.

While there are competing explanations for the increasing viability of niche products (for example, the fact that online markets and consolidations in the retail sector allow firms to serve a large audience than in the past), our account suggests that firms operating in markets in which traditional demographic data is less useful as a means of determining preferences benefit more from this trend than other markets. Empirical research examining the extent to which the rise of niche products differs between markets would therefore be a useful way to analyse the use of data in different industries hosted on online platforms.

7 Conclusion

Platforms collecting increasingly detailed data on consumer demographics and online behaviour not only poses a threat to consumers when a firm is a monopoly. Such data can be used selectively to increase the aggregate profits of firms on a platform by providing firms with sufficiently informative data such that less popular firms increase their price in segments of the population in which they are relatively popular, while at the same time not being too informative such that firms drop their prices aggressively in order to increase their sales in segments in which they are not popular.

The platform can further increase its profits by providing asymmetric data to the firms, as doing so potentially induces the less popular firm to set a price which only attracts buyers of its own type in a segment in which they are relatively popular, while not allowing the other firm to set a relatively low price in that segment. As big data collection techniques become more sophisticated, the potential for such asymmetric segmentations to increase profits will necessarily increase as well.

However, the fact that providing too precise a signal of consumer type to both firms is costly to the platform implies that how useful such big data will be depends on the distribution of consumers within a given market. We find that the platform has an incentive to reveal more granular data in markets where consumer preference is less polarised than in markets where broad demographic information is a relatively precise signal of consumer type. In a market where the niche firm is particularly unpopular, the platform is more likely to be able to reveal more precise information about consumer demographics and behaviour without creating too much competition than they are in markets where the type distribution is more even.

Consumers concerned solely with the prices they observe would prefer for both firms to be given complete information about their preferences. Platforms do not have an incentive to perfectly inform firms, instead preferring to provide partial information in order to decrease competition. A policy that constrains the platform's ability to selectively reveal the information it holds will increase consumer surplus if that data is sufficiently informative.

Of course, such a policy tool would not be cost-free in this context: the more information platforms provide to firms, the more there is potential for privacy concerns. Policies, like GDPR, which are designed to defend consumer privacy on online platforms may increase the ability of the platform to segment markets by centralising data collection, which results in firms operating on the platform having less information relative to the platform itself, allowing the latter to release the data selectively in the manner described here.

Appendix

Proof of Proposition 3

As $\rho_1 > \frac{1}{2}$, any equilibrium in the no-information hypergraph, it must be the case that $p_1^* \geq p_2^*$. Hence, the best response curves of the game can be summarised: $p_1^* = \frac{1+p_2^*}{2}$, $p_2^* = \frac{\rho_2+(1-\rho_2)(p_1^*)}{2(1-\rho_2)}$.

Now, consider the equilibrium in the complete information hypergraph. In such a case, if $b_i \in \mathcal{V}_j$ then: $p_j^c = \frac{1+p_2^*}{2}$ and $p_k^c = \frac{(1-\rho_2)(p_j^*)}{2}$. It is clear that $p_k^c < p_2^*$ in this case, and so $p_j^c < p_1^*$. It follows immediately that for the set of prices observed by any $b_i \in \mathcal{V}_1$, are both lower in the complete information hypergraph than in the no-information hypergraph.

It is possible (though by no means definite) that for $b_i \in \mathcal{V}_2$ s_2's price increases when there is complete information compared to the no-information case; that is it might be that $p_2^* < p_j^c$, and this in turn may increase profits on these buyers. However, as shown above $p_k^c < p_2^*$ and $p_j^c < p_1^*$, which implies that an increase in profits from a type 2 firm is outweighed by the decrease in profits from a type 1 firm. As $\rho_1 > \frac{1}{2}$, it follows that $H_n > H_c$.

Proof of Theorem 1

Let $E_k, E_k^{'} \in H_{\mathcal{S}}$, denote two edges in a symmetrically informative hypergraph generated by the same segment, $s_k \in \mathcal{S}$ and where $f_i \in E_k$ and $f_j \in E_k^{'}$. As $\mathcal{S}$ is an informative dominance-preserving segmentation, it must be that $p_{1k}^* > p_{2k'}^*$ for any segment $s_k \in \mathcal{S}$.

When $p_{1k'}^* > p_{2k}^*$, p_{ik}^* is only a function of $\alpha(v = 1|s_k)$ in so far as a change in $\alpha(v = 2|s_k)$ affects $p_{jk'}^*$ which follows from the fact that the reaction function of f_1 in s_k is such that $p_{1k'}^* = \frac{1+p_{2k}^*}{2}$, which is not directly affected by the value of $\alpha(v = 1|s_k)$. Meanwhile, f_2's reaction function is $p_{2k'}^* = \frac{\alpha_k+(1-\alpha_k)(p_{1k}^*{}')}{2(1-\alpha_k)}$, where $\alpha_i = \alpha(v = 2|E_i)$,and hence $p_{1k'}^* = \frac{(1-\alpha_k)+\frac{1}{2}\alpha_k}{\frac{3}{2}(1-\alpha_k)}$.

As the pair of prices, $p_{ik}^*, p_{jk'}^*$ are not directly affected by the prices in any other edges of $H_{\mathcal{S}}$, and $H_{\mathcal{S}}$ is symmetric, it follows that we can write $\pi_k(\alpha_k|s_k) := \pi_{1k'}(\boldsymbol{p_1^*},\boldsymbol{p_2^*}; \alpha_k) + \pi_{2k}(\boldsymbol{p_i^*},\boldsymbol{p_j^*}; \alpha_k)$. Given the above expressions for equilibrium prices and the profit function, $\pi_k(\alpha_k|s_k)$ is convex in α_k for $\alpha_k \in (0, \frac{1}{2})$. Then $\pi_P(\boldsymbol{\alpha}|\mathcal{S}) = \sum_k^{|\mathcal{S}|} \pi_k(\alpha_k|s_k)$ is convex in $\boldsymbol{\alpha}$, a vector whose kth component is α_k, when $\alpha_k \in (0, \frac{1}{2})$, which holds here by the fact that $\mathcal{S}$ is dominance-preserving.

Now consider $\mathcal{S}'$, which is also an informative, dominance-preserving segmentation but less informative than $\mathcal{S}$. It is well known (see Blackwell, 1953) that if a posterior probability distribution, μ, is more informative in the Blackwell sense than another, ϑ, with the same mean, then μ is a mean-preserving spread of ϑ. Hence, if $\mathcal{S}$ is more informative than $\mathcal{S}'$, then the distribution of type 2s it generates is a mean-preserving spread of the equivalent distribution generated by $\mathcal{S}'$. As $\pi_P(\boldsymbol{\alpha}|\mathcal{S})$ is convex, it follows that $\pi_P(\boldsymbol{\alpha}|\mathcal{S}) > \pi_P(\boldsymbol{\alpha}|\mathcal{S}')$ and hence $H_{\mathcal{S}} \succ H_{\mathcal{S}'}$.

Proof of Proposition 4

Let $\boldsymbol{p_i'}$ denote the price vector of f_i generated by H'. Suppose first that $\boldsymbol{p_j'} = \boldsymbol{p_j^*}$, and hence the price set by f_j in E_t is equal for both hypergraphs. Consider the set of edges, Φ_s, which are subsets of the targeted edge, $E_s \in H$ generated by taking the creation of the refinement of H, H'.

As per Proposition 1, H' is more informative than H, and hence $\exists E_k \in \Phi_s$, which is targeted by f_i when $\boldsymbol{p_j'} = \boldsymbol{p_j^*}$. Given $\boldsymbol{p_j'} = \boldsymbol{p_j^*}$ by assumption, p_{ik}^*, the equilibrium price for E_k is equal to p_{is}^*. This follows from the fact that, conditional on f_i setting a price, $p_{is} > p_{jt}$, the solution to the maximisation problem $\max_{p_{is}} \alpha(\upsilon = i|E_s)(1-\chi(p_{is}-p_{jt}))p_{is}$ is independent of the conditional distribution generated by the segmentation chosen by the platform owner. Hence, f_i has no incentive to change its price unless f_j changes its price in this case. If every edge in Φ_k is targeted, then no firm has an incentive to change its price, and therefore $\pi_P(\boldsymbol{p_i'},\boldsymbol{p_j'}) = \pi_P(\boldsymbol{p_i^*},\boldsymbol{p_j^*})$.

However, if there is an edge, $E_r \in \Phi_s$ which is not targeted, then f_i by definition sets a price, p_{ir}' in response to $\boldsymbol{p_j^*}$ which is lower than p_{is}^*. The existence of such an edge is possible given that H' is more informative than H. As prices are locally strategic complements in this setting, this fall in prices would result in a reduction in p_{jt}, further reducing the prices set by f_i in any edge, $E \in \Phi_s$. Therefore, $H \succ H'$.

Proof of Theorem 2

Suppose that f_1 receives no information and f_2 is shown a segmentation, $\mathcal{S}$, which is such that there are two segments. One segment is such that $\alpha(v = 2|s_1) = \delta > \rho_2$ and the other, s_2, contains the remaining buyers.

Now, consider the value of δ, $\hat{\delta} < \frac{1}{2}$ where f_2 is indifferent between setting a price $p'_{21} > p'_{1n}$ and deviating to set a price some price, p''_{21} where $p'_{1n} \geq p''_{21}$. To see such a threshold value exists, note that $\hat{\delta}$ satisfies the following condition:

$$\hat{\delta} p'_{21}(1 - p'_{21} + p'_{1n}) = p''_{2k}[\hat{\delta} + (1 - \hat{\delta})(1 - p''_{21} + p'_{1n})].$$

For a pair of edges generated by a segmentation s_i with $\alpha(v = 2|s_i) := \alpha_i$, the symmetric hypergraph generates the following prices: $p^S_{1i} = \frac{\frac{1}{3}(2-\alpha_i)}{1-\alpha_i}$ and $p^S_{2i} = \frac{\alpha_i+(1-\alpha_i)p^S_1}{2(1-\alpha_i)}$ for all $s_i \in \mathcal{S}$. On the other hand, H generates the price $p^H_{2i} = \frac{1+p'_{1n}}{2}$ if $p'_{2i} \geq p'_{1n}$ and $p_{2i} = \frac{\alpha_i+(1-\alpha_i)p'_{1n}}{2(1-\alpha_i)}$ otherwise and:

$$p'_{1n} = \frac{\frac{1}{3}(1 + \rho_1)}{\rho_2 + \gamma},$$

where γ is the number of type 1s in s_2. When $\alpha_1 = \frac{1}{2}$, $p^H_{2i} = \frac{1}{2} + \frac{1}{2} p'_{1n}$, and since $p^H_{2i} \geq p'_{1n}$, it is immediately follows that $\hat{\delta} \in (0, \frac{1}{2})$. It is also useful to note that as $p'_{1n} = \frac{\frac{1}{3}(2-\rho_2)}{1+\rho_2(1-\frac{1}{\delta})}$, increasing δ has the effect of increasing $\delta p'_{21}(1 - p'_{21} + p'_{1n}) - p''_{2k}[\delta + (1 - \delta)(1 - p''_{21} + p'_{1n})$, and hence increases relative payoff

Now, we show that there exists values of α_1 where $\alpha_1 \geq \hat{\delta}$ (and so f_2 strictly prefers to target s_1) and $p'_{1n} \geq p^S_{21}$,which immediately implies that $H \succ H_{\mathcal{S}}$. In this setting,

$\gamma = \rho_1 - |s_1| + \rho_2$. As $|s_1| = \frac{\rho_2}{\delta}$, $\rho_2 + \gamma = 1 + \rho_2(1 - \frac{1}{\delta})$. Let $\tilde{\delta} \in (0, \frac{1}{2})$ denote the value of δ where $p^{'}_{1n} = p^S_{21}$. Then the following expression holds:

$$3\tilde{\delta}^2 - \tilde{\delta}(1 - \rho_2) - \rho_2 = 0.$$

It follows that $\tilde{\delta} = \frac{1}{6}[1 - \rho_2 + \sqrt{(1-\rho_2)^2 + 12\rho_2}]$. For all $\rho_2 \in (0, \frac{1}{2})$ the following inequality holds:

$$\tilde{\delta} p^{'}_{21}(1 - p^{'}_{21} + p^{'}_{1n}) > p^{''}_{2k}[\tilde{\delta} + (1 - \tilde{\delta})(1 - p^{''}_{21} + p^{'}_{1n})].$$

As the left hand-side of this inequality is increasing in $\tilde{\delta}$, it follows that $\tilde{\delta} > \hat{\delta}$, and hence when $\delta \in (\hat{\delta}, \tilde{\delta})$, $p^{'}_{1n} \geq p^S_{21}$. This immediately implies that $p^{'}_{21} > p^S_{11}$.

Proof of Theorem 3

Let $E_k, E^{'}_k \in H_{\mathcal{A}}$, denote two edges in a symmetric hypergraph generated by the same segment, $s_k \in \mathcal{S}_A$ and where $f_1 \in E_k$ and $f_2 \in E^{'}_k$. As $\mathcal{S}_A$ is dominance-preserving, it follows that $p^*_{1k} > p^*_{2k'}$ for $H_{\mathcal{A}}$. Suppose that H is the hypergraph that is generated by the same segmentation as H_A in $\mathcal{M}_B$ and hence is identical to H_A under the conditions in $\mathcal{M}_A$. Then, $p^*_{1k} > p^*_{2k'}$ for all edge pairs $E_k, E^{'}_k \in H$.

By the proof of Theorem 1, when $p^*_{1k} > p^*_{2k'}$, p^*_{1k} is only a function of $\alpha(v = i|s_k)$ in so far as a change in $\alpha(v = 2|s_k)$ affects $p^*_{jk'}$. Given the functional form of the demand function, equilibrium prices are convex in $\alpha(v = j|s_k, \mathcal{M}_A)$. As the pair of prices, $p^*_{ik}, p^*_{jk'}$ are not directly affected by the prices in any other edges of H, it follows that we can write $\pi_k(\alpha_{jk}|\mathcal{M}_A, s_k) := \pi_{ik}(\boldsymbol{p^*_i}, \boldsymbol{p^*_j}; \alpha_{jk}) + \pi_{jk'}(\boldsymbol{p^*_i}, \boldsymbol{p^*_j}; \alpha_{jk})$, where $\alpha_{jk} = \alpha(v = j|s_k, \mathcal{M}_A)$ and $\pi_k(\alpha_{jk}|s_k, \mathcal{M}_A)$ is a convex function. Then $\pi_P(\boldsymbol{\alpha}_2|\mathcal{M}_A, \mathcal{S}_A) = \sum_k^{|\mathcal{S}_A|} \pi_k(\alpha_{jk}|\mathcal{M}_A, s_k)$ is convex in $\boldsymbol{\alpha}_2$, a vector whose kth component is α_{jk}.

If $\mathcal{S}_A$ is more informative than $\mathcal{S}$, then it is a mean-preserving spread of $\mathcal{S}$. By the definition of informativeness, if $\mathcal{S}_A$ is dominance-preserving, then$\mathcal{S}^{'}$ is also dominance-preserving: if there was some segment, s_t, where $\alpha(\upsilon = j|s_k) > \frac{1}{2}$, then it could not be that $\mathcal{S}_A$ is more informative than $\mathcal{S}^{'}$. As $\pi_P(\boldsymbol{\alpha}_2|\mathcal{M}_A, \mathcal{S}_A)$ is convex in $\boldsymbol{\alpha}_2$, it follows that $H_A \succ H^{'}$, the hypergraph generated by the segmentation $(\mathcal{S}, \mathcal{S})$.

If $\mathcal{S}_A$ and $\mathcal{S}$ are dominance-preserving under the conditions in $\mathcal{M}_A$, then the same is true in $\mathcal{M}_B$, by the definition of polarisation. Similarly, the polarisation assumption implies that $\mathcal{S}_A$ is still a mean-preserving spread of $\mathcal{S}$ under the conditions in $\mathcal{M}_B$.

If H_A and H_B are symmetrically informative such that both firms are shown the dominance preserving segmentations $\mathcal{S}_A$ and $\mathcal{S}_B$ respectively then it follows that $\mathcal{S}_A$ cannot be more informative than $\mathcal{S}_B$. If this were the case, then H would be strictly preferred to H_B for the market $\mathcal{M}_B$, and hence $H_B \notin \Lambda_B^*$.

Proof of Proposition 5

If $\mathcal{S}_A$ is dominance-preserving in $\mathcal{M}_A$, it is also dominance-preserving in $\mathcal{M}_B$. This follows because $\alpha(\upsilon = 2|s_k, \mathcal{M}_A) > \alpha(\upsilon = 2|s_k, \mathcal{M}_B)$ and $\alpha(\upsilon = 2|s_k, \mathcal{M}_A) < \frac{1}{2}$ by definition. Furthermore, if $\mathcal{S}$ less informative than $\mathcal{S}_A$ then $\mathcal{S}$ is dominance-preserving in both $\mathcal{M}_A$ and $\mathcal{M}_B$. It follows from the proof of Theorems 1 and 3 that any $H_{\mathcal{S}}$ that is less informative than H_A is dispreferred to H_A under the conditions in both $\mathcal{M}_A$ and $\mathcal{M}_B$. It follows immediately that H_B cannot be less informative than H_A if it is dominance-preserving and symmetrically informative.

References

[1] Acquisti, A. Inducing customers to try new goods. *Review of Industrial Organization 44*, 2 (2014), 131–146.

[2] Ali, S. N., Lewis, G., and Vasserman, S. Voluntary disclosure and personalized pricing. *Proceedings of the 21st ACM Conference on Economics and Computation* (2020).

[3] Armstrong, M. Recent Developments in the Economics of Price Discrimination. Industrial Organization 0511004, University Library of Munich, Germany, Nov. 2005.

[4] Armstrong, M., and Zhou, J. Consumer information and the limits to competition. Economics Series Working Papers 888, Nov. 2019.

[5] Belleflamme, P., and Vergote, W. Monopoly price discrimination and privacy: The hidden cost of hiding. *Economics Letters 149*, C (2016), 141–144.

[6] Bergemann, D., and Morris, S. Information design, bayesian persuasion, and bayes correlated equilibrium. *American Economic Review 106*, 5 (May 2016), 586–91.

[7] Bester, H., and Petrakis, E. Coupons and oligopolistic price discrimination. *International Journal of Industrial Organization 14*, 2 (1996), 227–242.

[8] Blackwell, D. Equivalent Comparisons of Experiments. *The Annals of Mathematical Statistics 24*, 2 (1953), 265 – 272.

[9] Bounie, D., Dubus, A., and Waelbroeck, P. Selling strategic information in digital competitive markets. *The RAND Journal of Economics 52*, 2 (2021), 283–313.

[10] Cecere, G., Le Guel, F., Manant, M., and Soulié, N. *The Economics of Privacy.* Palgrave Macmillan UK, London, 2017, pp. 1–11.

[11] Charlson, G. Searching for results: Optimal platform design in a network setting. Tech. rep., 2020.

[12] Chen, J., Elliott, M., and Koh, A. Capability Accumulation and Conglomeratization in the Information Age. Cambridge Working Papers in Economics 2069, Faculty of Economics, University of Cambridge, July 2020.

[13] Chen, Y. Paying customers to switch. *Journal of Economics & Management Strategy 6*, 4 (1997), 877–897.

[14] DziubiÅski, M., and Goyal, S. How do you defend a network? *Theoretical Economics 12*, 1 (2017), 331–376.

[15] Elliott, M., Galeotti, A., and Koh, A. Market segmentation through information. Cambridge working papers in economics, Faculty of Economics, University of Cambridge, 2021.

[16] Esteves, R.-B., and Cerqueira, S. Behavior-based pricing under imperfectly informed consumers. *Information Economics and Policy 40*, C (2017), 60–70.

[17] Esteves, R.-B., and Resende, J. Personalized pricing and advertising: Who are the winners? *International Journal of Industrial Organization 63*, C (2019), 239–282.

[18] Fudenberg, D., and Tirole, J. Customer poaching and brand switching. *The RAND Journal of Economics 31*, 4 (2000), 634–657.

[19] Johnson, J. P., and Myatt, D. P. On the simple economics of advertising, marketing, and product design. *American Economic Review 96*, 3 (June 2006), 756–784.

[20] Kamenica, E., and Gentzkow, M. Bayesian persuasion. *American Economic Review 101*, 6 (October 2011), 2590–2615.

[21] Myerson, R. Conference structures and fair allocation rules. *Int J Game Theory* (1980).

[22] Neiman, B., and Vavra, J. The rise of niche consumption. Becker Friedman Institute For Economics 2019-102, University of Chicago, https://papers.ssrn.com/sol3/papers.cfm?abstract$_i d = 3432536, 2020$.

[23] Novshek, W., and Sonnenschein, H. Fulfilled expectations cournot duopoly with information acquisition and release. *Bell Journal of Economics 13*, 1 (1982), 214–218.

[24] Raith, M. A general model of information sharing in oligopoly. *Journal of Economic Theory 71*, 1 (1996), 260–288.

[25] Roesler, A.-K., and Szentes, B. Buyer-optimal learning and monopoly pricing. *American Economic Review 107*, 7 (July 2017), 2072–80.

[26] Shiller, B. First degree price discrimination using big data. Working Papers 58, Brandeis University, Department of Economics and International Businesss School, 2014.

[27] Slikker, M., Dutta, B., van den Nouweland, A., and Tijs, S. Potential maximizers and network formation. *Mathematical Social Sciences 39*, 1 (2000), 55–70.

[28] Tao, H., Bhuiyan, M. Z. A., Rahman, M. A., Wang, G., Wang, T., Ahmed, M. M., and Li, J. Economic perspective analysis of protecting big data security and privacy. *Future Generation Computer Systems 98* (2019), 660–671.

[29] Thisse, J., and Vives, X. On the strategic choice of spatial price policy. *American Economic Review 78*, 1 (1988), 122–37.

[30] Townley, C., Morrison, E., and Yeung, K. Big Data and Personalized Price Discrimination in EU Competition Law. *Yearbook of European Law 36* (12 2017), 683–748.

[31] van den Nouweland, C., Borm, P., and Tijs, S. Allocation rules for hypergraph communication situations. *International Journal of Game Theory 20*, 3 (1992), 255–268. Pagination: 13.

[32] Vives, X. Aggregation of information in large cournot markets. *Econometrica 56*, 4 (1988), 851–876.