Essays in Industrial Organisation

by

Andrew Rhodes

Nuffield College

Submitted to the University of Oxford

in Hilary Term 2010, in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

Essays in Industrial Organisation

by

Andrew Rhodes

Nuffield College

Submitted to the University of Oxford
in Hilary Term 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis presents four largely independent essays on industrial organisation. The first three essays examine how search and switching costs distort competitive markets, whilst the fourth essay studies how firms themselves might eliminate competition by agreeing to fix the market price. The key insight from the first two essays is that by choosing to pay a search cost, a consumer reveals to a retailer some private information about their product valuations. In this context the first essay re-examines the well-known theoretical Diamond Paradox in which markets completely break down if firms sell only a single product. I demonstrate that multiproduct retailers offer an elegant and realistic way of overcoming this Paradox, and then apply the model to supermarkets. In particular, this essay provides new insights into why convenience stores charge high prices and why grocery stores use selected loss-leaders. The second essay looks at internet search and therefore focuses on the special case in which search frictions become very small. It seeks to explain why retailers pay so much for online advertising, when consumers can easily click on whichever links they like. I show that if consumers have prior information about products, their search behaviour is limited but very informative about their preferences. This places prominent firms in a privileged position, and makes them substantially more profitable. The third essay provides a simple model in which small switching costs are pro-competitive and beneficial to consumers. This challenges the conventional wisdom, but also argues that switching costs should be less of a policy priority than search costs. The final essay examines a game in which two firms bargain over a collusive price. It is shown that entry into such a market may make collusion easier, and may increase price.

"it is often true that only by going too far can we find out how far we can go"

I would like to begin by thanking my supervisor, Paul Klemperer, for his guidance and encouragement over the past four years. I have learnt many things from him which I am sure will stand me in good stead far into the future. Especial thanks are also due to David Myatt and Chris Wilson, who have helped greatly shape the way I think about many of the problems discussed here.

I have also received helpful comments from many others, including but not limited to Mark Armstrong, Heski Bar-Isaac, Meg Meyer, Charles Roddie, Sandro Shelegia, Guiying Wu and Jidong Zhou.

Financial assistance from both the Economic and Social Research Council (PTA-031-2005-00119) and the Royal Economic Society is also gratefully acknowledged.

Finally, and most importantly, my biggest debt is to my family, especially my mum and late nan, without whose love and support it would never have been possible to come this far.

Declaration

I declare that this thesis represents my own work, and that (with the exception of Chapter 5, which is based on my M.Phil. thesis) none of it has already been accepted, or is concurrently being submitted, for any degree or diploma or certificate or other qualification in this University or elsewhere.

Signature ..... Date .....

Word Count: 59130 words including bibliography and appendices (270 words per page, multiplied by 219 pages)

Contents

1 Introduction **6**

2 Multiproduct Pricing and the Diamond Paradox **9**

2.1 Introduction 10

2.2 Assumptions 15

2.3 Characterising Equilibrium Prices 17

2.4 Comparative Statics 23

2.5 Discussion 35

2.6 Extensions 38

2.7 Conclusion 44

2.8 Appendix 46

3 Non-Random Search on the Internet **52**

3.1 Introduction 53

3.2 Assumptions 56

3.3 Equilibrium Pricing Behaviour 58

3.4 Increasing the Degree of Prominence 73

3.5 Discussion 78

3.6 Conclusion 83

3.7 Appendix 84

4 Small Switching Costs are Pro-Competitive **92**

4.1 Introduction 93

4.2 Literature 95

4.3 Assumptions 98

4.4 Partial Discrimination 100

4.5 Full Discrimination 111

4.6 No Discrimination 119

4.7 Who Benefits from Price Discrimination? 133

4.8 Conclusion and Policy Discussion 138

4.9 Appendix 140

5 Price Collusion when Firms have Different Costs **159**

5.1 Introduction 160

5.2 Related Literature 163

5.3 The Model 165

5.4 The Pricing Supergame 166

5.5 Alternating Offers Bargaining Game 169

5.6 Limit of Frictionless Bargaining 185

5.7 Entry into Collusive Markets 188

5.8 Conclusion 200

5.9 Appendix 201

Bibliography **210**

List of Figures

2-1 The firm attracts consumers with relatively high valuations 19

2-2 Shoppers and Diamond consumers 24

2-3 The effect of product range on price 28

2-4 The effect of demand shocks on price 30

2-5 Niche versus eclectic product selections 41

3-1 How prices are affected by changes in $\tilde{s}$ 75

4-1 'Ripoffs-then-bargains' Pricing 110

4-2 Prices under No Discrimination 131

4-3 Maximum Equilibrium Prices under No Discrimination 132

5-1 A low-cost entrant makes collusion more sustainable 192

5-2 Relatively low-cost entry facilitates collusion 193

5-3 Collusion when c_h is small 199

5-4 Collusion when c_h is medium 199

5-5 Collusion when c_h is large 199

Chapter 1

Introduction

The core of the thesis is four short essays on industrial organisation. Chapters 2 and 3 study the impact of search frictions on the high street and internet respectively, whilst Chapter 4 focuses on switching costs. Chapter 5 is a shortened version of my M.Phil. thesis.

My first essay (Chapter 2) analyses the pricing and advertising behaviour of a multiproduct monopolist, whose prices can only be learnt by a consumer upon payment of a positive search cost. The Chapter begins by proposing a solution to the well-known theoretical Diamond Paradox [20]. The Paradox states that when consumers have unit demand and strictly positive search costs, the market entirely breaks down and no trade occurs. However I demonstrate that multiproduct retailers overcome this Paradox in a particularly elegant and yet realistic manner. The key insight is that by choosing to pay a search cost, a consumer reveals to a retailer some private information about his product

valuations. Retailers end up exploiting this information, and in a single-product world the market breaks down. With multiple products, less can be inferred about a consumer's valuations, and trade can occur. The Chapter then uses this insight to explain some aspects of supermarket behaviour. In particular, I show that the model can help us understand why small convenience stores charge high prices, understand how selected loss-leaders can commit a retailer to a store-wide low-price image, and also explain why products are often marked down when in high demand.

My second essay (Chapter 3) attempts to reconcile the apparent ease of online search with the prevalence of sponsored advertisements. Small changes in the way products are displayed on a webpage appear to have large effects on demand. Consequently firms collectively pay large sums of money to advertise their products in prominent positions, such as at the top of a Google webpage. However it is not clear why prominence should matter so much - because free 'organic' search results are displayed nearby, and consumers can click wherever they like. The key insight from the model, much like in Chapter 2, is that a consumer's clicks reveal information about his preferences. Therefore if consumers have some prior information about the set of available products, small frictions cause all prices to double even when search is random. I then demonstrate that when some links are more prominent, the consumers who click them reveal less match information. Prominent firms therefore charge lower prices and earn more profit - but interestingly do not appear to earn more profit per-click. This suggests that standard per-click auctions might fare badly in certain markets. The essay also demonstrates that better search

algorithms - by successfully matching more consumers with their preferred product - increase prices and therefore hurt consumers.

My third essay (Chapter 4) demonstrates that small switching costs are probably pro-competitive and good for consumers. On the one hand switching costs may lead to high prices because they encourage firms to exploit past customers, but on the other hand may lead to low prices as firms compete aggressively for new customers who can then be exploited later on. Klemperer [45] (and most other authors) argues that on balance switching costs are anti-competitive and allow firms to sustain high profits. I provide a model which I argue to be more realistic than many others, and come to the opposite conclusion. The essay also demonstrates that prices may follow an unusual 'ripoffs-then-bargains' shape, and that even an incumbent monopolist facing a brand new entrant, may find it more worthwhile to compete for future market share than harvest its old customers.

My final essay (Chapter 5) is a shortened version of my M.Phil. thesis. Two firms with heterogeneous marginal costs meet up and bargain over a collusive price. The bargaining game has the unusual features that the two parties' preferences are not completely opposed, and also any bargaining solution must respect incentive compatibility in the subsequent pricing supergame. These differences lead to interesting properties of the equilibrium and some unusual comparative statics. The essay also argues that entry into a market may help improve the chances of collusion, and may also place upward pressure on collusive prices.

Chapter 2

Multiproduct Pricing and the Diamond Paradox

Abstract: I show that multiproduct firms suffer less than single-product firms from the "Diamond Paradox". Equilibrium prices are high because rational consumers understand that visiting a store exposes them to a hold-up problem when they have search costs. However a store with more products attracts more consumers with low valuations, and therefore charges lower prices. Advertising a few products at low prices enables the firm to credibly commit to low prices across the rest of the store. "Loss-leading" can therefore be optimal. If demands are subject to random shocks, prices are shown to move counter cyclically.

2.1 Introduction

Typically retailers only advertise the prices of a few of their products (and many others do no advertising at all). Consequently consumers usually learn the price of a particular product only once they have arrived at the retailer. Despite all of the marketing and online price comparisons available to supermarket shoppers, even in that sector most price information is probably still gathered instore.[1] However shopping (especially for basics) can be time-consuming, so visiting stores and learning about prices is not the easy and costless activity which many papers assume it to be. I therefore present a simple model which captures this important feature. Consumers must pay a (small) shopping cost in order to travel to a multiproduct monopolist and discover its prices.

The first contribution of the paper is to show that multiproduct firms can resolve the Diamond Paradox [20]. The exact details of this Paradox depend upon whether a consumer has unit or downward-sloping demand. Most of the literature (implicitly or explicitly) assumes unit demand, and so does this paper. Suppose firms sell a single product, and consumers must pay a cost $s > 0$ to learn any retailer's price. If consumers expect each retailer to charge p^E, they only visit a firm if they value the product more than $p^E + s$. So once somebody has turned up to a store, they will buy the product there provided its actual price is less than $p^E + s$ (the consumer still gets positive surplus, and

[1] Simester [65] suggested that (in 1995) a typical supermarket stocked $25,000$ products and advertised 200 of them. Large UK supermarkets stock $30,000$ different products ('The rise and rise of Lidl Britain', The Daily Telegraph, September 10^{th} 2008). Tesco.com allows a comparison of some $10,000$ prices with its main rivals, but may be slightly out of date; searching the website is time-consuming; smaller outlets are excluded; and relatively few people may use such sites often.

it is not worthwhile to pay another s and visit another retailer). Therefore firms always charge more than consumers expected. The only equilibrium outcome of the model is for consumers to expect very high prices, nobody to visit any retailer, and no trade to occur.

Unsurprisingly, many papers have suggested possible ways to overcome this 'no trade' result. Possible resolutions include advertising (Wernerfelt [82]), product differentiation and unknown match values (Anderson and Renault [1], Konishi and Sandfort [46]), as well as repeated interaction between consumers and producers (Bagwell and Ramey [8]). In Stahl [67], some consumers are 'shoppers' - meaning they have zero search costs and learn every price in the market. The remaining consumers pay $s > 0$ for each price quotation that they learn. Firms cut prices to win business from the shoppers, and hence trade occurs.[2]

In this paper I interpret the Diamond Paradox as occurring because of a 'sample selection problem'. Only consumers with relatively high valuations find it worthwhile to go shopping. Therefore retailers always want to charge high prices, which ultimately causes the market to collapse. Multiproduct retailers can overcome the paradox, because they attract a broader selection of consumers. Somebody with a low valuation on one product may visit a store because they have a high valuation on something else. Consequently consumers within the store are more representative of the population. This

[2]Burdett and Judd [14] present a similar idea. All consumers are ex ante identical, but the number of prices learnt upon paying s is stochastic and sometimes exceeds 1. Hence some consumers have better information, which encourages firms to cut price.

restrains the firm's incentives to surprise consumers with high prices, and makes equilibria with trade possible. Although every consumer has the same strictly positive search cost, some consumers behave *as if* they are shoppers in the sense of Stahl.

The model also provides several insights into firm behaviour. It predicts that a store with a broader product selection should charge lower prices but earn more profit on each good. A simple rationale is also provided for why low advertised prices on a few products can credibly signal low prices on the rest of the store's (unadvertised) products. Prices are also predicted to be countercyclical.

In a recent survey, 86% of American consumers said they thought larger stores (with broader product selections) charge lower prices.[3] Consistent with this, Hoch *et al* [37] find that larger grocery stores have more elastic demand curves. There is also abundant anecdotal evidence that bigger retailers tend to charge less. For example although the major UK supermarkets use national pricing, this does not apply to their smaller stores, which are typically more expensive.[4][5] The main explanations for this are costs and convenience. Larger stores may enjoy economies of scale and buyer power, which they pass on in lower prices. Smaller (convenience) stores may also attract time-poor but cash-rich consumers. The model in this paper presents a different interpretation. A

[3] 'American Consumer Institute Survey Finds Consumers Prefer Shopping at Larger Stores and Wholesale Clubs', PR Newswire, 13th April 2006. A Canadian price survey suggests that store size is indeed the biggest determinant of price ('Bigger grocery stores less expensive: study', The Gazette (Montreal), April 5th 2007).

[4] In 2006 Tesco was accused of charging up to 25% more in some smaller stores ('MPs demand tough retail regulator', The Sunday Telegraph, 12th February 2006).

[5] At the same time it is undoubtedly true that there does exist significant dispersion in prices across retailers (Sorensen [66]).

retailer always prices to a select sample of relatively high-valuation consumers. However a store with more products attracts more consumers, and these consumers have on average lower valuations. Hence a larger retailer finds it optimal to charge lower prices. In particular, if one were to exogenously add extra products into a retailer's store, it would find it optimal to reduce all of its prices - even prices of unrelated products. Indeed the recent drive towards one-stop shopping[6] can then be interpreted as an attempt by firms to commit to low prices across their whole product range.

Most retailers either do no price advertising, or they advertise a small number of products at very low prices. Occasionally products are even sold below cost, with examples ranging from books[7] to realtime share prices[8]. Supermarkets are also well-known for using loss-leaders to boost store traffic.[9] This raises an interesting question: what effect does price advertising have on overall price levels? There are two main viewpoints. On the one hand, retailers may increase the prices of unadvertised products, to make up lost margin on items that are sold cheaply. On the other hand, advertising a few products may enable efficient retailers to signal their low costs (and general low price level).[10] The papers by Lal and Matutes [48] and Simester [65] capture these ideas.

In Lal and Matutes [48] consumers have unit demands and identical willingnesses to

[6] See for example 'Post Office Stocks up for Fight', Marketing, September 14th 2005. Supermarkets also now stock a range of previously non-core products, such as electricals and garden furniture.

[7] 'Harry Potter 7: wizard of a loss leader', The Vancouver Sun, July 7th 2007.

[8] 'British operator to offer stock data free on Net', The Vancouver Sun, July 15th 1999

[9] See Appendix 5.6 of the Competition Commission's 2008 Groceries Final Report. All of the major supermarkets admitted to using loss-leaders, with these accounting for on average 3% of revenue. Increasing store traffic was one stated reason.

[10] See for example John Fingleton's argument in 'Remove the shackles from retail and distribution sectors', The Irish Times, May 30th 2003.

pay. One product is unadvertised and consumers pay their reservation price for it. Another product is advertised, and used solely to compete for store traffic. In Simester [65] firms also stock two products, and advertise the price of only one. Consumers have identical and downward-sloping demand for the unadvertised good. A firm's production cost is correlated across products, and more efficient firms charge less for their unadvertised product. They may signal this by advertising a low price on the other product.

The crucial assumption in my paper is that consumers have heterogeneous willingnesses to pay. This implies that a retailer receives only a select sample of consumers who have relatively high product valuations. When a product is advertised at a low price, this draws new consumers into the store. These new visitors must have relatively low valuations for other products (otherwise they would have visited even without the advertising). The retailer optimally cuts prices on unadvertised product lines in an attempt to win more business from the new consumers. Hence the model provides an intuitive explanation for how low advertised prices on a few products can successfully signal a low store-wide price image. Loss-leadership pricing is also shown to sometimes be optimal.[11]

It is well-known that retailers tend to mark down items during periods of high general demand - such as weekends and holidays. Further, when individual items are in high demand, they are often advertised at low prices. (See Warner and Barsky [80] and MacDonald [51]) One possible explanation for countercyclical pricing is Rotemberg and

[11] Empirical evidence is ambiguous. Milyo and Waldfogel [54] study the lifting of a ban on alcohol price advertising. They find little effect on overall alcohol prices, and no effect on the prices of unadvertised alcohol products. However their sample is small, and they cite other (cross-sectional) papers which do find that advertising results in lower prices. Collins *et al* [19] present some evidence that banning loss-leaders results in high prices.

Saloner's [61] model of collusion. Alternatively, Warner and Barsky argue that during weekends, consumers visit more shops and so become more responsive to prices. Bils [11] argues that new customers may have less brand attachment. When lots of new consumers enter the market, it is optimal to cut price and persuade them to try the product. In my model, weekends mean that more consumers are interested in buying a product. New consumers visit the store, and they tend to have low valuations on other products. Hence the retailer again has more incentives to charge lower prices across the whole of its product range.

The rest of the paper proceeds as follows. Section 2.2 sets out the main assumptions. In Section 2.3 I solve the model and demonstrate how a multiproduct retailer can overcome the Diamond Paradox. Comparative statics results are provided in Section 2.4, whilst Sections 2.5 and 2.6 discuss the results and provide possible extensions.

2.2 Assumptions

There is a single firm that produces n goods, indexed by $j = 1, 2, ..., n$, at zero marginal cost. The products are neither substitutes nor complements, and consumers demand at most one unit of each. Consumer valuations for the n goods are denoted $(v_1, v_2, ..., v_n)$. For each consumer, v_l and v_k are independent whenever $l \neq k$. Each v_j is drawn from $[a_j, b_j]$ (where $b_j > 0$) using a distribution function $F_j(v_j)$ (with corresponding density $f_j(v_j)$). $f_j(v_j)$ is strictly positive, continuously differentiable, and logconcave. This ensures that the hazard rate $\frac{f_j(p)}{1-F_j(p)}$ is increasing, and holds for many common

distributions.[12] In the textbook zero-search-cost model, each good's profit function is strictly quasiconcave and has a unique maximiser $p_j^m = \arg\max p\left[1 - F_j\left(p\right)\right]$.

The monopolist may advertise A prices, where A is strictly less than n and could be zero. The firm is legally obliged to honour all advertised prices. Consumers use advertisements to form rational *expectations* about all prices, denoted $\left(p_1^E, p_2^E, ...p_n^E\right)$. Consumers must pay a shopping cost $s > 0$ to visit the store[13]. Once incurred, this cost is sunk. Prior to visiting the store, consumers know their valuations $(v_1, v_2, ..., v_n)$. Using their *expectations* about price, they turn up if and only if they expect to earn a surplus greater than the shopping cost. After they have arrived at the store, consumers learn *actual* prices $(p_1, p_2, ...p_n)$ and make their purchases. All parties are risk neutral and rational. The move order can be summarised as:

1. The monopolist chooses which goods (if any) to advertise, and at what prices. It then chooses (but does not disclose) the prices of the remaining goods

2. Consumers form expectations about all prices, and decide whether or not to visit the store

3. Consumers who decided to turn up then learn actual prices, and make purchase decisions

[12] See Bagnoli and Bergstrom [7].

[13] I use the terms 'search cost' and 'shopping cost' interchangeably. s is a search cost because it must be paid to learn prices. But it is also a shopping cost because it must be paid even by a consumer who only buys advertised products.

2.3 Characterising Equilibrium Prices

Equilibrium in this model has three requirements. First, consumers only visit the store if they expect to earn enough surplus to cover the shopping cost (given their valuations and expected prices). Second, actual prices must maximise firm profit (given expected prices and therefore given the types of consumers who visit the store). Third, expected prices must equal actual prices (rational expectations).

Write the demand for an unadvertised good (call it good 1) as:

$$D_1 = \int_{p_1}^{b_1} f(v_1) \Pr\left(\sum_{j=1}^{n} \max\left(v_j - p_j^E, 0\right) \geq s \right) dv_1 \tag{2.1}$$

A consumer buys good 1 if (a) he values it more than its *actual* price, and (b) he turns up to the store. Turning up is only worthwhile if total expected surplus $\sum_{j=1}^{n} \max\left(v_j - p_j^E, 0\right)$ exceeds the shopping cost s. Through this turn-up decision, demand for any one product depends upon the expected prices of all goods. Nevertheless conditional upon visiting the store, demand for a product depends only upon its actual price. This is because goods are neither substitutes nor complements.[14]

In equilibrium, the firm must maximise its profit by setting p_1 equal to p_1^E. Given any vector of expected prices $\left(p_2^E, ..., p_n^E\right)$, there is at most one p_1^E where this is true, and

[14] Notice that when p_1 hits $p_1^E + s$, demand kinks and becomes more responsive to changes in price. Hence one can rationalise the finding that sometimes consumers are loss-averse with respect to price changes.

it satisfies the following interior[15] first order condition[16]

$$D_1|_{p_1=p_1^E} - p_1^E f_1\left(p_1^E\right) \underbrace{\Pr\left(\sum_{j=2}^{n} \max\left(v_j - p_j^E, 0\right) \geq s\right)}_{\text{Probability a consumer with } v_1=p_1^E \text{ turns up}} = 0 \tag{2.2}$$

To understand (2.2), consider a small increase in p_1 above the expected level p_1^E. The firm *gains* revenue on existing consumers, who have mass equal to demand, but *loses* revenue on consumers who stop buying good 1 following the price rise. Each consumer who stops buying the good (a) has a marginal valuation for it, and (b) visits the store. Part (a) explains the $f\left(p_1^E\right)$ term, whilst part (b) explains the term $\Pr\left(\sum_{j=2}^{n} \max\left(v_j - p_j^E, 0\right) \geq s\right)$. Intuitively, since the consumer is marginal for good 1 but has turned up, he must have expected surplus of s or more on the other $n-1$ goods. Figure 2-1 provides a graphical intuition when $n = 2$. Only consumers in the shaded region $A + B + C$ visit the store, so the firm gets an 'adverse selection' of high-valuation types. When $p_1 = p_1^E$, demand for product 1 equals $B + C$. Small changes in p_1 around p_1^E only affect the behaviour

[15] An interior solution is one that satisfies $p_1^E > a_1$, whilst a corner solution satisfies $p_1^E = a_1$. A necessary (but not sufficient) condition for a corner solution is that $p_1^m = a_1$. At a corner solution, the lefthand side of (2.2) is (almost always) strictly negative.

To understand why, suppose that $p_1^m = a_1$. When $s = 0$, profit is (almost always) strictly decreasing in p_1 around $p_1 = a_1$; intuitively each marginal consumer is very valuable, so small price increases (which lose marginal consumers) are bad for profit. The firm therefore prefers to sell good 1 to everybody. Now suppose that $s > 0$, consumers expect a price $p_1^E = a_1$, and $\Pr\left(\sum_{j=2}^{n} \max\left(v_j - p_j^E, 0\right) \geq s\right)$ is sufficiently large. Since each marginal consumer is very valuable, it can again be the case that profit is strictly decreasing in p_1 around $p_1 = a_1$, and that the firm prefers to sell good 1 to everybody who turns up to its store.

[16] Condition (2.2) only checks for small deviations in p_1 around p_1^E, and may have multiple solutions, but at most one is compatible with profit being globally maximised at $p_1 = p_1^E$. Profit is quasiconcave in p_1 if $\Pr\left(\sum_{j=2}^{n} \max\left(v_j - p_j^E, 0\right) \geq s\right) \geq \frac{1}{2}$ or if the standard monopoly profit function $p\left[1 - F_1\left(p\right)\right]$ is concave. See Claims 1 and 2 in the Appendix.

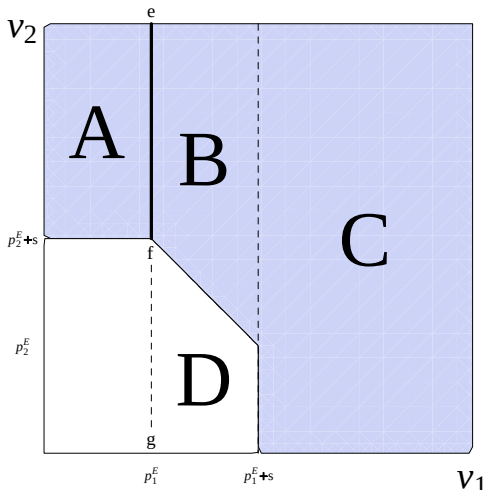


Figure 2-1: The firm attracts consumers with relatively high valuations

of consumers on the thick line $e\ f$. All other marginal consumers have $v_2 < p_2^E + s$, so they do not visit the store and do not observe changes in p_1.

Figure 2-1 also helps draw a parallel with the Diamond Paradox. If only good 1 is sold and its price is unadvertised, the firm faces a strong sample selection problem, because only high-valuation consumers in region C turn up. In particular, no consumer with $v_1 = p_1^E$ visits the store. Hence the pricing condition (2.2) is only satisfied when $D_1|_{p_1=p_1^E} = 0$ - i.e. no trade.[17] A key result of the paper is that a multiproduct firm can attract marginal consumers, and thus make trade possible:

Proposition 1 *When n is sufficiently large, there exist equilibria in which the firm optimises, consumer expectations are fulfilled, and trade occurs*

When $n = 1$ trade breaks down because no marginal consumer ever visits the store.

[17] If $n = 1$ but $s = 0$, demand for good 1 is $B + C + D$ and all marginal consumers (on line $e\ f\ g$) visit the store. Hence (2.2) simplifies to $1 - F_1\left(p_1^E\right) - p_1^E f_1\left(p_1^E\right) \leq 0$ - the standard monopoly first order condition.

By contrast when n is sufficiently large, many consumers visit the store, and many of these have marginal valuations for each product. Demand curves become sufficiently sensitive to small price changes, so the firm is deterred from surprising consumers with price increases. Of course if there is no advertising, no-trade Diamond equilibria exist as well. (If consumers expect very high prices, nobody visits the store, so the firm is indifferent between all prices and is happy to charge whatever consumers expected) However the Diamond outcome is not a very compelling prediction of possible play. For example, suppose the monopolist must incur an ϵ entry cost. The firm would never enter if it expected a zero-profit Diamond outcome; by entering the market, the firm signals to consumers that it expects to play a non-Diamond equilibrium. Discussion of equilibrium multiplicity and (where required) selection is left till the next section.

This naturally raises the question of what 'sufficiently large' means in practice. Whilst this will depend very much on the specific distributions of product valuations, it is easy to provide some illustrative examples - found in the accompanying table. The Table shows, for a specific distribution of valuations and shopping cost, what a sufficient number of products would be in order to avoid the Paradox.

	Shopping Cost			
	$s \approx 0$	$s = 1$	$s = 2$	$s = 5$
$N(0,1)$ valuations	7	22	40	92
$N(1,1)$ valuations	4	9	15	33

To interpret the numbers, suppose that a product were free - with $N(0,1)$ valuations and $s = 2$, less than 2.5% of the population would find it worthwhile to go shopping.

Hence in this context $s = 2$ is rather a large search cost. But even so, 40 products would be sufficient to avoid the Diamond Paradox. Shifting up the distribution of valuations also helps, because it somewhat reduces the retailer's optimal price. Therefore although difficult to fully interpret, these numbers do suggest that the Diamond Paradox is relatively easy to overcome.

Before concluding this section, rewrite the interior pricing condition (2.2) as

$$\underbrace{\Pr\left(\sum_{j=2}^{n} t_j \geq s\right)}_{\text{Shoppers}} \underbrace{\left[1 - F_1\left(p_1^E\right) - p_1^E f_1\left(p_1^E\right)\right]}_{\text{Standard first order condition}} + \underbrace{\Pr\left(\sum_{j=1}^{n} t_j \geq s > \sum_{j=2}^{n} t_j\right)}_{\text{Diamond consumers}} = 0 \qquad (2.3)$$

where $t_j = \max\left(v_j - p_j^E, 0\right)$ is expected surplus on good j. Pricing is affected by the behaviour of three different groups. Consumers with $\sum_{j=1}^{n} t_j < s$ do not expect to cover the shopping cost, and therefore do not visit the store. They do not observe the actual price p_1 and so their demand for good 1 remains at zero. Consumers with $\sum_{j=1}^{n} t_j \geq s$ do visit the store, and subdivide into two groups:

- **Shoppers** for product 1 have $\sum_{j=2}^{n} t_j \geq s$
- **Diamond consumers** for product 1 have $\sum_{j=1}^{n} t_j \geq s > \sum_{j=2}^{n} t_j$

Shoppers for product 1 turn up at the store regardless of their valuation for the product. Consequently they reveal no information about their v_1, which continues to be distributed on $[a_1, b_1]$ with the usual density $f_1(v_1)$. The profit earned on these consumers is simply $p_1\left[1 - F_1(p_1)\right]$ (the same as in a standard zero-search-cost monopoly problem),

so small changes in p_1 around p_1^E affect profits by $1 - F_1\left(p_1^E\right) - p_1^E f_1\left(p_1^E\right)$. The term 'shoppers for product 1' is used to describe these consumers because they act *as if* they have no shopping cost when it comes to buying good 1. In contrast, Diamond consumers for product 1 have $v_1 - p_1^E > 0$. They would all continue to buy product 1 even if the firm increased p_1 slightly above p_1^E. Their demand is locally perfectly inelastic - just like the consumers in a standard Diamond problem. *Therefore the multigood problem with positive shopping cost is really an average of the standard monopoly and Diamond problems.*

Lemma 2 *When trade occurs, $p_j^E \geq p_j^m$ for each unadvertised good j*

A small increase in p_1 above p_1^E always gains extra revenue on Diamond consumers; in equilibrium, this must be balanced by losses on shoppers. Consequently the equilibrium price of product 1 exceeds p_1^m. Since the firm's pricing problem is an average of the standard monopoly and Diamond problems, in general p_1^E strictly exceeds p_1^m.

Corollary 3 *The firm prefers equilibria with lower unadvertised prices*

Interestingly, both consumers and the firm itself benefit from a small reduction in unadvertised prices. Intuitively, a decrease in p_1^E benefits the firm in two ways. Firstly, it lowers the price of product 1 closer to p_1^m and therefore makes the product more profitable. Secondly, the price decrease draws more consumers into the store, expanding demand and therefore profits on other products in the store.

2.4 Comparative Statics

I first solve the model for small shopping costs. I then show that the results generalise to arbitrary s, provided two natural conditions are imposed.

Comparative static results are best understood using sample selection and shoppers/Diamond consumers. A shopper for product 1 has $\sum_{j=2}^{n} t_j \geq s$ whilst a Diamond consumer for the same product has $\sum_{j=1}^{n} t_j \geq s > \sum_{j=2}^{n} t_j$. Figure 2-2 illustrates this in $\left(v_1, \sum_{j=2}^{n} t_j\right)$ space. The distribution of $\sum_{j=2}^{n} t_j$ has mass at 0, which I represent using an area. It is useful to split the Figure into three regions and analyse a shift up in the distribution of $\sum_{j=2}^{n} t_j$. *Region I*: consumers with $v_1 \leq p_1^E$ only visit the store if $\sum_{j=2}^{n} t_j \geq s$. So if $\sum_{j=2}^{n} t_j$ increases, more consumers visit the store and become shoppers for product 1. *Region III*: consumers with $v_1 \geq p_1^E + s$ definitely visit the store, and are shoppers or Diamond consumers for product 1 depending upon whether $\sum_{j=2}^{n} t_j$ exceeds s. Clearly then, as $\sum_{j=2}^{n} t_j$ increases, the number of shoppers increases and the number of Diamond consumers falls. *Region II* deals with $v_1 \in \left(p_1^E, p_1^E + s\right)$. An increase in $\sum_{j=2}^{n} t_j$ converts some Diamond consumers to shoppers. It also attracts new people to the store, some of whom become shoppers and others who become Diamond consumers. Therefore in Regions I and III, an increase in $\sum_{j=2}^{n} t_j$ raises the ratio of shoppers to Diamond consumers, whilst Region II is ambiguous. I now propose two methods to overcome this ambiguity.

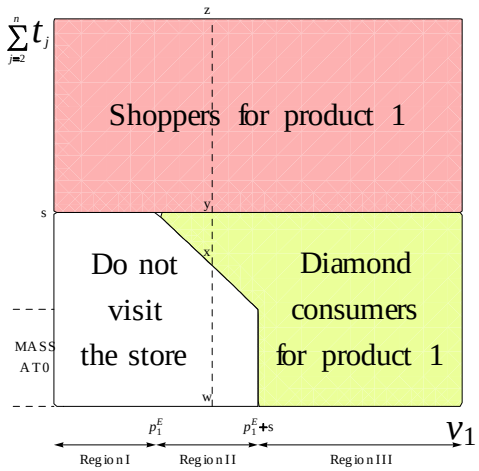


Figure 2-2: Shoppers and Diamond consumers

2.4.1 Small Shopping Costs

I begin by focusing on the special case $s \to 0_+$ (the shopping cost is strictly positive but arbitrarily small). The mass of consumers with $v_1 \in \left(p_1^E, p_1^E + s\right)$ - Region II in Figure 2-2 - becomes vanishingly small. Referring to the above discussion, it is clear that a shift up in the distribution of $\sum_{j=2}^{n} t_j$ will increase the probability of being a shopper and decrease the probability of being a Diamond consumer for product 1.[18] The interior pricing condition (2.3) can then be simplified to

$$|\epsilon_1| = \frac{1}{1 - \prod_{j=2}^{n} F_j\left(p_j^E\right)} \tag{2.4}$$

where ϵ_1 is the price elasticity of demand in a standard (zero-search-cost) monopoly problem. In equilibrium $|\epsilon_1|$ is equated with the inverse probability of a marginal consumer

[18] Limit probabilities are $1 - \prod_{j=2}^{n} F_j\left(p_j^E\right)$ and $\left[1 - F\left(p_1^E\right)\right] \prod_{j=2}^{n} F_j\left(p_j^E\right)$ respectively.

visiting the store. If the latter is equal to 1, $|\epsilon_1| = 1$ and $p_1^E = p_1^m$. (Note that $|\epsilon_1|$ is strictly increasing in p_1^E)

Small search frictions induce complementarities in the way different products are priced. Suppose for example that consumers suddenly expect p_2^E to be higher. The distribution of $\sum_{j=2}^{n} t_j$ is shifted down, so the probability of being a shopper for product 1 decreases whilst the probability of being a Diamond consumer increases. With more Diamond consumers to exploit through price increases, and fewer shoppers to tempt with price cuts, the firm wants to charge more than originally expected for good 1. Consequently its equilibrium price increases - even though the search friction is minimal and the two products are independent in both use and valuation.

Complementarity in pricing decisions opens up the possibility of multiple equilibria. If the firm commits to $\left(p_2^E, ..., p_n^E\right)$ via advertising, there is a single p_1^E that solves (2.4) and therefore (subject to profit being quasiconcave) a unique equilibrium. Multiplicity becomes possible whenever there is more than one unadvertised product. To illustrate, suppose that goods 1 and 2 are unadvertised and that any other prices are fixed (perhaps by advertising). If p_1^E is expected to be 'high', product 2 has few shoppers and therefore p_2^E is also expected to be high - in which case product 1 has few shoppers and the expectation that p_1^E is high can be justified. At the same time if p_1^E is expected to be 'low', p_2^E is also expected to be low which in turn can justify the original expectation about p_1^E. Expectations can therefore be self-reinforcing, as the following (striking but atypical) example demonstrates:

Example 4 *No advertising, and valuations are iid with* $F(v) = \ln \frac{v}{a}$ *(*$\ln \frac{b}{a} = 1$ *and* $p^m = a$*).*[19] *Equilibria are symmetric*

When $n = 2$*, any* $p^E \in [a, b]$ *is an equilibrium*

When $n \geq 3$*, the only equilibria are* a *and* b

Complementarity ensures existence of a Pareto Dominant equilibrium. Hold constant any advertised prices and look for equilibrium vectors of unadvertised prices. Since the equilibrium price of product j is an increasing function of the expected prices of all other goods, Tarski's Fixed Point Theorem guarantees the existence of a lowest price vector. By Corollary 3, this is also be the Pareto Dominant equilibrium.

Assumption: *Agents play the Pareto Dominant equilibrium*

This assumption is probably less severe than it might first appear. There always exists a region of fixed entry costs such that the firm only enters if it expects to play the Pareto Dominant equilibrium. Hence by Forwards Induction Logic, entering coordinates expectations on the low-price equilibrium. Often there is one low-price equilibrium and several high-price (low-profit) equilibria, so the region of entry costs for which this works can be quite large. Burning money to artificially increase fixed costs, as well as cheap-talk in advertisements, can also help focus beliefs.[20]

[19] *Although* $f(v) = v^{-1}$ *is not logconcave,* $p[1 - F(p)]$ *is concave.*

[20] In equilibrium there would be no need to actually burn the money.

Forwards Induction can work in other ways too. For example, stocking n goods may give a unique equilibrium, but n' goods give both a low and a high equilibrium. Choosing n goods may be best if with n' the high equilibrium is played, but choosing n' goods may be best if the low equilibrium is played. By choosing n' goods the firm signals it expects the low equilibrium.

Proposition 5 *A store with more products charges lower prices*

Precisely, if advertised prices are held constant and n is increased, unadvertised products become cheaper.[21] When a new product is introduced, some additional consumers are attracted to the store, and they act as shoppers for each existing product. Also some consumers who were previously turning up but were reliant on a particular product, are not any longer. Hence each existing product receives more shoppers and less Diamond consumers. The firm wants to lower its unadvertised prices and expand output sold to these shoppers. So consumer expectations adjust downwards and equilibrium prices are lower.[22] As n grows large, the price of each product gets close to its standard monopoly level and the firm can almost extract all monopoly rents. See also Armstrong [3].

When prices are unadvertised and distributions are identical, non-Diamond equilibria exist if and only if n exceeds a threshold $\tilde{n}$. In the case of $U[a,b]$ valuations, $\tilde{n} = 2 - \frac{a}{b}$ and any non-Diamond equilibrium is unique. Four examples are shown in Figure 2-3 (in each case $p^m = \frac{1}{2}$). As n increases the firm attracts a broader mix of consumers who have lower valuations. Hence the sample selection problem is less severe and equilibrium price decreases. With $U[0,1]$ valuations, price is very close to $1/2$ even when $n = 10$. Intuitively if consumers expect every product to have price $\frac{1}{2}$, the probability they turn

[21] Some restrictions are needed to ensure that profit is quasiconcave in the price of the new product. It is sufficient that valuations for the new product have a similar distribution to an existing product, and/or the conditions given earlier hold.

[22] There is an an analogy with Keynesian multipliers. Suppose initial price expectations are q_0. Given expectations q_0, with the new product the firm wants to charge $q_1 < q_0$. If consumers then expect q_1, shoppers become yet more numerous and the firm wants to charge prices $q_2 < q_1$ instead. Consumers can then expect prices q_2 and so forth. The decreasing sequence $\{q_t\}_{t=0}^{\infty}$ converges to the rational expectations equilibrium. The initial increase in n gets successively multiplied up into lower and lower prices.

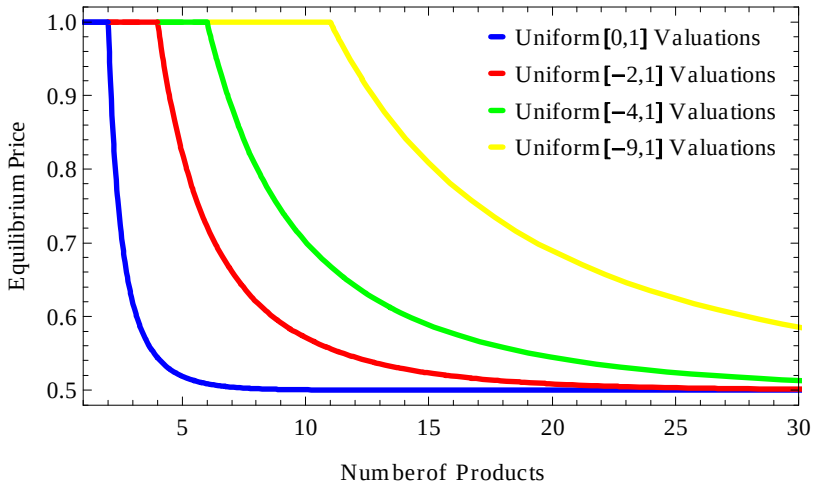


Figure 2-3: The effect of product range on price

up is $1-\frac{1}{2}^{10} \approx 1$. There is almost no sample selection problem, so the firm prices almost as if there were no shopping cost. However as a is reduced, the tail of the distribution is pulled down and there are more low-valuation consumers for each product. Given any expected price, fewer people visit the store and the firm faces a more adverse selection. Therefore as a is reduced, equilibrium price increase. I now reintroduce advertising and consider its impact on equilibrium.

Proposition 6 *If the monopolist reduces an advertised price, the prices of unadvertised products fall as well*

Lower advertised prices - just like increased product variety - attract a broader mix of consumers to the store. The firm therefore faces a weaker sample selection problem, and responds by charging less on each unadvertised product. The implicit assumption behind the Proposition is that a firm must (for legal or other reasons) not renege on any

price that it has committed to via advertising. Conditional upon that assumption, an advertised price on one product is informative about the prices of everything else. In particular, the model gives a theoretical justification behind the idea that low advertised prices on some (it need not be many) products can be very effective at building a 'low-price' image on the rest of the store's product portfolio.

I now consider the effect of product-specific demand shocks. A simple way to model these is to place the following structure on consumer valuations:

$$v_j : \begin{cases} = -\infty & \text{with probability } 1-\alpha_j < 1 \\ \overset{iid}{\sim} [a,b] \text{ via } F(v) & \text{with probability } \alpha_j \end{cases}$$

An increase in α_j means that product j is more popular, and ceteris paribus is demanded more often by consumers. Within a convenience store, for example, certain products like milk will have a higher α than other products such as shampoo. When any advertised prices are held constant, the following result obtains:

Proposition 7 *A positive demand shock on one product reduces all unadvertised prices. Within the store, more popular unadvertised products have higher prices*

Consider the first part of the Proposition. Higher demand for any one product brings extra people into the store, especially those with low valuations. Hence the firm faces a less adverse selection, and charges lower prices across the store. Prices are therefore countercyclical. Now consider the second part of the Proposition. A very popular product attracts many consumers to the store and therefore provides less-

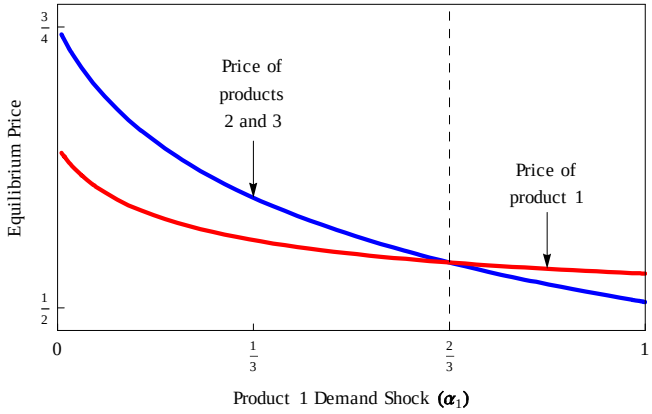


Figure 2-4: The effect of demand shocks on price

popular products with many shoppers. However these less-popular products are not an important factor in consumers' decision to visit the store, and therefore do not provide many shoppers for the popular products. This gives the firm a greater incentive to surprise consumers with price cuts on unpopular products (to attract the relatively large number of shoppers) and price rises on popular products (to attract the relatively large number of Diamond consumers).

Figure 2-4 illustrates graphically the effect of demand shocks. In the example there are three unadvertised products, $F(v) = 2v - 1$, and $\alpha_2 = \alpha_3 = \frac{2}{3}$. When α_1 increases above $\frac{2}{3}$, product 1 becomes the most important factor affecting whether consumers visit the store, and hence it becomes relatively more expensive. But the prices of all goods are strictly decreasing in α_1.

To summarise, increased product variety, low-price adverts, and positive demand shocks, all attract a broader mix of consumers to the store. This eases the firm's sample selection problem and reduces unadvertised prices. Although store traffic increases in

each case, it is the broader mix of consumers that is crucial. For example suppose with probability $1-\beta$ a consumer has $v_j = -\infty \forall j$, and with probability $\beta > 0$ product valuations are *iid* on interval $[a, b]$. Increases in β raise store traffic, but the mix is unchanged - new visitors have the same joint distribution over valuations as old visitors. Consequently the firm's pricing incentives are unchanged, and there is no effect on equilibrium prices.

2.4.2 General Shopping Costs

It is intuitive that the comparative statics results in the previous section continue to hold whenever s is 'sufficiently small'.[23] These results also hold for a general s in simple environments where the firm only stocks two products. They also extend to arbitrary numbers of products provided that two natural conditions are imposed on the distribution of consumer surplus.

Proposition 8 *Consider an arbitrary s*

- ***Product range*** *Assume identical distributions and no advertising*

  Equilibrium price falls when n is increased from 2 to 3

- ***Advertising*** *Assume $n = 2$ and one product is advertised*

  The equilibrium unadvertised price increases in the advertised price

[23] Region II in Figure 2-2 remains small, so increases in $\sum_{j=2}^{n} t_j$ still increase the ratio of shoppers to Diamond consumers for product 1.

- ***Demand shocks*** *When $n = 2$, Proposition 7 holds*

I now show how the comparative statics results can be extended beyond the two-product environment considered in Proposition 8. Let $T_{m,p}$ be the total expected surplus from m products when facing expected price vector $p = \left(p_1^E, p_2^E, ..., p_n^E\right)$. The two conditions on $T_{m,p}$ are as follows:

- **C1** Surplus decreases in price in the sense of hazard rate dominance

$$\frac{\Pr\left(T_{m,p} \geq z\right)}{\Pr\left(T_{m,q} \geq z\right)} \text{ increases in } z,\ z \in (0, s) \text{ and } q > p$$

- **C2** Surplus increases in m in the sense of hazard rate dominance

$$\frac{\Pr\left(T_{m,p} \geq z\right)}{\Pr\left(T_{m-1,p} \geq z\right)} \text{ increases in } z,\ z \in (0, s)$$

$T_{m,p}$ is always increasing in m and decreasing in p in the sense of first order stochastic dominance (FOSD). Conditions C1 and C2 are stronger, and require that conditional distributions satisfy FOSD as well. Hazard rate dominance is a common assumption in problems with multidimensional types.[24]

C1 ensures that pricing problems are complementary. To illustrate this, suppose that the vector $\left(p_2^E, p_3^E, ..., p_n^E\right)$ increases. Product 1 becomes more expensive if and only if its ratio of Diamond consumers to shoppers increases (since the firm then has more

[24]Hazard rate dominance is weaker than (and implied by) the monotone likelihood ratio property. The latter is commonly used. See for example Milgrom [53].

incentives to surprise consumers with a high price). A sufficient condition for this is that region II in Figure 2-2 is well-behaved. Take a typical locus of points in this region $w\ x\ y\ z$. We require that the probability of being on $x\ y$ relative to the probability of being on $y\ z$, is increasing in the price vector $\left(p_2^E, p_3^E, ..., p_n^E\right)$. This is the same as requiring that $\Pr\left(\sum_{j=2}^n t_j \geq z\right) / \Pr\left(\sum_{j=2}^n t_j \geq s\right)$ is increasing in price, or that C1 holds. This ensures that a Pareto Dominant equilibrium exists, and that unadvertised prices are positively associated with all other prices.

C2 governs comparative statics in product range, search cost, and demand shocks. Assume either C1 is satisfied or that distributions are identical (in which case a Pareto Dominant equilibrium exists). C2 says that conditional on visiting the store (having $\sum_{j=1}^n t_j \geq s$), a consumer is less likely to be a shopper for a product (have $\sum_{j \neq k} t_j \geq s$) when s is larger. Equivalently when the search cost increases, Diamond consumers become more numerous relative to shoppers. It follows immediately that when s increases, the firm has incentives to charge more on each product than was previously expected - and hence equilibrium unadvertised prices increase. Stocking an additional product has the opposite effect. Some consumers expect to get positive surplus from the new product, and this effectively decreases their shopping cost. As a result shoppers become relatively more numerous, the firm has incentives to reduce prices, so in equilibrium unadvertised products become cheaper. A positive demand shock has the same effect: when α_j increases, extra surplus is randomly allocated across consumers so their effective search cost falls and the firm wants to charge less on each product.

2.4.3 Continuity of Equilibrium around $s = 0$

Equilibrium prices need not be discontinuous around $s = 0$, but usually are. If the firm stocks a *single* product and s increases from 0 to something slightly positive, the equilibrium price jumps up from p^m to b. This discontinuity does not happen in a multiproduct context if there exists some good l which is advertised at a price $p_l < a_l$. This is because every consumer visits the store to buy l and is a shopper for all unadvertised products - which are consequently priced at their respective standard monopoly levels. Ruling this out, a necessary condition for avoiding discontinuity is that there exists an unadvertised product j with $p_j^m = a_j$. A sufficient condition is that there exists a second unadvertised product k that also has $p_k^m = a_k$. The reasoning is best illustrated through an example. Suppose $n = 2$, $v_1 \sim U\left[\frac{2}{3}, 1\right]$ with $p_1^m = \frac{2}{3}$, and $v_2 \sim U\left[\frac{1}{3}, 1\right]$ with $p_2^m = \frac{1}{2}$. Let $s \to 0_+$ and look for an equilibrium in which $p_1^E \to \frac{2}{3}$ and $p_2^E \to \frac{1}{2}$. In the limit everybody visits the store to buy product 1 and is therefore a shopper for product 2 - rationalising the expectation $p_2^E \to p_2^m = \frac{1}{2}$. Only two-thirds of consumers (those with $v_2 > \frac{1}{2}$) are shoppers for product 1, the rest being Diamond consumers. Nevertheless using these probabilities, the lefthand side of pricing condition (2.3) is strictly negative when evaluated at $p_1^E = \frac{2}{3}$ so this is an equilibrium. Intuitively even marginal consumers are very valuable to the firm since the distribution is shifted up a lot relative to marginal cost. This explains why the firm strictly prefers to sell to everybody when $s = 0$. Even when $s \to 0_+$, provided there are sufficiently many valuable marginal consumers in the store, the firm still wishes to sell to them all.

2.5 Discussion

> "large volume operations create an impression of lower prices... lots of advertising, and a wide assortment... are the accepted cues for lower prices; a small store is considered the strongest indicator of high prices... [loss leaders are] associated with high volume operations." (Brown 1969)

This quotation from a consumer survey captures the essence of the model very well. It is often argued that retailers with high volumes enjoy lower costs, which they pass on in lower prices. This paper suggests another channel - sample selection. A retailer always receives relatively high-valuation consumers from the population. This selection problem is less severe whenever the store sells a broader product range, has high-demand goods, and uses low-price advertising. The model also has many interesting implications.

2.5.1 Informative advertising

Large retailers often advertise (very) low prices on a selection of their products. Simester [65] argues that low-price advertising can signal a low marginal cost, and hence signal low prices across the whole store. Lal and Matutes [48] argue instead that unadvertised prices are high and unrelated to advertised prices.

My model gives a different perspective - advertising is informative. I demonstrate that low-price advertising on a few products, can act as a credible signal of store-wide low prices. This happens even when products are independent in both use and valuation. A retailer should clearly never advertise a product at a price above its standard

monopoly level. Hence there is a natural dispersion between cheap advertised products and expensive unadvertised items. Nevertheless consumers are sophisticated. They use advertisements to make inferences about other prices, whilst recognising that advertised prices are typically much lower. Selling certain products below cost can also be a profitable strategy, though examples usually require large s and/or discrete value distributions

Example 9 *Two products with valuations iid* $U[0,1]$. *If* $s = \frac{4}{5}$ *and only one product is advertised, it is sold at a loss*

According to the model, products should be advertised at low prices when their demand is exogenously high. We saw in Proposition 7 that the most popular products are also the most expensive when they are unadvertised. Advertising them therefore has two benefits. Firstly, their price can be reduced and their profitability greatly increased. Secondly, because they are so popular, advertising them cheaply helps bring lots of new consumers to the store, which helps commit to significantly lower prices on the rest of the product range. There is lots of empirical evidence that products tend to fall in price when their demand is high (MacDonald [51], Warner and Barsky [80]). This might seem puzzling, but the model provides an intuitive rationale for this behaviour. The model also says that prices in general should be lower when aggregate demand is higher. Empirical evidence suggests this is also true - retailers seem to mark down items at weekends and during holiday seasons (Warner and Barsky [80]).

Kaul and Wittink [41] report that advertised products usually have more elastic de-

mand curves. The model presented in this paper provides a simple explanation. If product 1 is unadvertised, its demand is equal to $\int_{p_1}^{b_1} f_1(v_1) \Pr\left(\sum_{j=1}^{n} \max\left(v_j - p_j^E, 0\right) \geq s\right) dv_1$. If instead it is advertised, its demand equals

$$\int_{p_1}^{b_1} f_1(v_1) \Pr\left(\max(v_1 - p_1, 0) + \sum_{j=2}^{n} \max\left(v_j - p_j^E, 0\right) \geq s\right) dv_1$$

Demand for good 1 is more responsive to changes in p_1 when the price is advertised. The reason is that turnup decisions now depend upon p_1 rather than on just a fixed expectation p_1^E.

2.5.2 Product Line Spillovers

The model also suggests a different way of thinking about product range. Convenience retailers (as well as specialist/niche suppliers) are interpreted as being 'trapped' into charging unprofitably high prices. These retailers attract a small subset of consumers who tend to be interested in only a few items and therefore have very inelastic demand curves. Expecting very high prices, only consumers with especially high valuations turn up to the store. By contrast, stores with a broader product range and/or higher-demand items charge lower prices but earn proportionately higher profits (so doubling each α_j more than doubles profits). This works solely through demand elasticity and has nothing directly to do with demand, although the resulting price decreases do of course expand

demand.[25] Consequently introducing a new product brings both a direct benefit (its own profit) and an indirect one (higher profit on other items). Nevertheless (assuming introducing a new product has some fixed cost) it is unclear whether optimal product range increases or decreases when moving from $s = 0$ to $s \to 0_+$. On the one hand a new product brings indirect benefits and is therefore more valuable. But on the other hand prices are high and profit on the new product is therefore low, so it is less valuable. Either effect may dominate, depending upon the precise example. Nevertheless the model may partly explain the recent drive towards one-stop shopping.

2.6 Extensions

2.6.1 Correlation

It is simple to argue that no trade can occur in equilibrium if valuations are perfectly positively correlated and all prices are unadvertised. However advertising is very effective in the presence of correlation. To illustrate, suppose there are n products and $v_j = v$. v is drawn from $[a, b]$ and the standard monopoly price is $p^m > a$. If the firm advertises one product at price $p^m - s$, the other $n - 1$ (unadvertised) products cost p^m. This is an equilibrium because all consumers with $v \in [p^m, b]$ visit the store. When s is small, advertising is then a very effective way of committing to low prices.

[25] For instance the comparative statics results were easiest to prove when $s \to 0_+$. But in that case demand for one product was essentially totally independent of all other prices. The comparative statics in n worked only at the margin.

When valuations are affiliated and s is small, the comparative statics results in Section 2.4.1 continue to hold. Letting $s \to 0_+$, we can rewrite (2.4)

$$|\epsilon_1| = \frac{1}{\Pr\left(\sum_{j=2}^{n} \max\left(v_j - p_j^E, 0\right) \geq s \middle| v_1 = p_1^E\right)} \tag{2.4*}$$

Therefore fixing $\left(p_2^E, p_3^E, ..., p_n^E\right)$ there is a unique equilibrium p_1^E that is increasing in other expected prices. Adding new products (and/or advertising old ones at lower prices) brings in more consumers who are marginal for product 1. This again makes demand for product 1 more elastic and pushes down its equilibrium price.

It is interesting to consider whether the monopolist should choose a niche or an eclectic product range. Niche products are rated similarly (you either love or hate everything), but eclectic products can be valued very differently. To be concrete, suppose *eclectic* products have valuations independently and uniformly distributed on $[a, b]$. Further, consumer i's valuation for the *niche* product j is $v_{ij} = x_i + y_j$. x_i is a consumer-specific taste parameter that is $U[a, b]$; y_j is product-specific and is $U[-\epsilon, \epsilon]$ where ϵ is small but at least an order of magnitude larger than s.[26] I assume $\frac{b}{2} > a$ and look for symmetric equilibria.

Imagine the firm must choose between stocking $\bar{n}$ niche products or $\bar{n}$ eclectic products - which should it choose? If p_{ni} and p_{ec} denote equilibrium prices under the two

[26] Within the population, each niche product's valuation is approximately uniformly distributed on $[a, b]$. Since ϵ is small, the distribution is exactly uniform everywhere except very close to a and b.

arrangements, then it is simple to demonstrate that

$$\begin{aligned} \frac{1-F(p_{ni})}{p_{ni}f(p_{ni})} - 1 + \frac{1}{n} &= 0 \\ \frac{1-F(p_{ec})}{p_{ec}f(p_{ec})} - 1 + F(p_{ec})^{n-1} &= 0 \end{aligned}$$

(Note that as $n \to \infty$, price tends towards the standard monopoly price in both cases[27]) Sometimes an eclectic mix always delivers a lower price (and higher profit) regardless of $\bar{n}$. Otherwise (and this is usually the case) the optimal choice is to pick a niche product range when $\bar{n}$ is small, and an eclectic mix when $\bar{n}$ is large. This is illustrated in Figure 2-5 when valuations are uniformly distributed on $[0, 1]$. The crucial factor is always how many marginal consumers visit the store. When $\bar{n}$ is small and products are eclectic, few marginal consumers turn up - valuations for other products are dispersed and prices relatively high. But if products are niche, the marginal consumer for product 1 has a high x and therefore high valuations for other products - making it quite likely he will visit the store. Consequently when $\bar{n}$ is small, a niche selection delivers more marginal consumers and therefore lower price. As shown in the figure, this reverses when $\bar{n}$ is sufficiently large. Intuitively some consumers who are marginal for product 1 had quite a low x_i but a high y_1 draw. To persuade them to visit the store requires a large y_j draw on some other product j - and to achieve this requires many extra products to be added. On the other hand, with an eclectic selection, each new product is totally different and

[27] It may seem surprising that this happens when products are niche. As $n \to \infty$, many consumers do not visit the store. However they all have low valuations (for everything) so the firm never wants to sell to them anyway. The key is that as $n \to \infty$, all marginal consumers turn up.

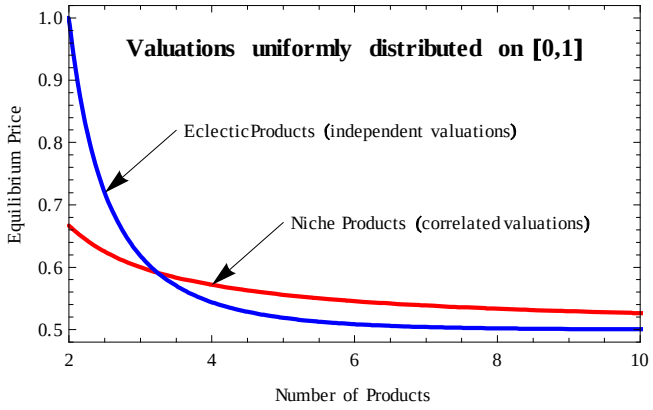


Figure 2-5: Niche versus eclectic product selections

valuations are completely random - therefore it is much easier to attract new marginal consumers.

2.6.2 Substitutes

Imagine a situation in which the retailer sells several products but they are all substitutes - meaning that a consumer wants to buy at most one of them. It is simple to argue that without advertising, only a no-trade Diamond equilibrium exists.

An advertised price on one product acts as a signal about the prices of other substitute products. To illustrate, suppose there are two goods with valuations independently and uniformly distributed on $[0, 1]$. Product 1 is unadvertised, and product 2 is advertised at price p_2^E. Suppose the firm charges slightly more for product 1 than consumers expect. Most consumers who intended to buy the product, still do. However some decide to buy product 2 instead. Clearly in equilibrium $p_1^E > p_2^E$ - if the reverse were true, the firm would benefit when people substituted from product 1 to product 2, and therefore

definitely charge more for good 1 than expected. There is a unique equilibrium, with

$$p_1^E = \frac{1 + 4p_2^E + s}{3}$$

As expected $\frac{\partial p_1^E}{\partial p_2^E} > 0$ - an increase in p_2^E both increases the demand for product 1 *and* makes it less costly when consumers substitute from good 1 and good 2. So the firm wants to charge more for product 1. One problem is that since p_1^E is high, profit is only quasiconcave if $p_2^E \leq \frac{1}{3} - s$ (otherwise p_1^E is so high that the firm may benefit from charging less than expected). If we let $s \to 0$ and advertise product 2 at a price of $\frac{1}{3}$, then the unadvertised product's equilibrium price is $\frac{7}{9}$. This large difference occurs because otherwise the firm would have a very strong incentive to surprise consumers and charge more for good 1 than they were expecting. Nevertheless advertising has commitment value. This brief extension captures in a stylised way Footnote 34 (and surrounding discussion) of Ellison and Ellison [23].

2.6.3 Shoppers

Up to now I have assumed that everybody has the same strictly positive shopping cost. I also showed that the monopolist's problem could be rewritten in such a way that some consumers acted *as if* they were shoppers and had zero shopping cost (in the sense of Stahl [67]). I now briefly consider what happens if some consumers have an exogenously given zero search cost.

Suppose that a fraction γ of consumers have $s \equiv 0$, whilst the remaining $1 - \gamma$ have

$s \to 0_+$. Then the analogue of (2.4) is

$$|\epsilon_1| = \frac{1}{1-(1-\gamma)\prod_{j=2}^{n} F_j\left(p_j^E\right)} \tag{2.4**}$$

Intuitively a consumer is only not marginal if she is not a shopper in the exogenous sense (happens with probability $1-\gamma$) and also not a shopper endogenously (which happens with probability $\prod_{j=2}^{n} F_j\left(p_j^E\right)$). It is then apparent that the usual comparative statics continue to hold. In addition, if the proportion of shoppers increases, more of the population visits the store and the firm can infer less about valuations, and hence charges lower prices. These results generalise provided that the non-shoppers' s is not too large.

2.6.4 Competition

Fuller investigation of the effects of competition is deferred to later. In this short section I briefly demonstrate that in the absence of advertising, competition does not influence pricing decisions.

Suppose there are several retailers, and they each stock the same n products. There is no price advertising, and look for a symmetric equilibrium in which each firm charges the same prices. An individual retailer could make small adjustments to prices and not cause any consumers to search another retailer. Hence at the margin, a competitive firm has the same pricing incentives as the monopolist did in the earlier part of this paper. In particular there is always an equilibrium in which retailers charge the same prices

as a multiproduct monopolist would when facing consumers with search cost s. This obviously has the flavour of the original Diamond [20] result.

2.7 Conclusion

Multiproduct firms help resolve the Diamond paradox, but the essential intuition about prices being high persists. Only consumers with high valuations turn up, and the store exploits them by raising prices. Consumers understand this, so in equilibrium price is high, and profits low. A firm with a broader and frequently-demanded product range has less incentive to hold consumers up, and so charges lower prices. The model therefore provides an intuitive explanation for why retailers have embraced one-stop shopping by expanding into (previously) non-core activities. Further, low-price advertising on a few products acts as a commitment to charging low prices across the whole store. Hence the model also provides a novel explanation for why firms sometimes use loss-leaders. In addition, prices may move countercyclically and products should be advertised at low prices when their demand is high.

Three interesting extensions are competition, bundling, and mergers. Firstly, by extending the set-up to competition, there is the potential to help explain why prices of even everyday items are priced very differently by different retailers. Secondly, the model also suggests that retailers would have an incentive to commit not to bundle products - since bundling generally reduces variation in valuations and therefore worsens the sample selection problem. And finally the model suggests that mergers of two unrelated firms

could result in falling prices, even if there are no cost synergies. These are all interesting areas which I hope to explore soon.

2.8 Appendix

Proof of equation (2.2): Let $T = \sum_{j=2}^{n} \max\left(v_j - p_j^E, 0\right)$. Differentiating $p_1 D_1$ with respect to p_1 gives $D_1 - p_1 f_1(p_1) \Pr\left(T + \max\left(p_1 - p_1^E, 0\right) \geq s\right)$ $(\star)$. This is continuous in p_1 for $p_1 < p_1^E + s$. Setting $p_1 = p_1^E$ gives (**2.2**). Now check that $p_1 = p_1^E$ is globally optimal. Charging $p_1^E + s$ (weakly) dominates any higher price, since when $p_1 \geq p_1^E + s$, $p_1 D_1 = p_1 [1 - F_1(p_1)]$ which is decreasing in p_1. Moreover charging a_1 (weakly) dominates any lower price. Therefore for an *interior* solution, we need to check that $p_1 D_1$ is strictly increasing in p_1 for $p_1 \in \left[a_1, p_1^E\right)$ and strictly decreasing in p_1 for $p_1 \in \left(p_1^E, p_1^E + s\right]$. For a *corner* solution, we need only check that $p_1 D_1$ is strictly decreasing in p_1 for $p_1 \in \left(p_1^E, p_1^E + s\right]$. There are two ways. (1) rewrite $(\star)$ as $[D_1 - p_1 f_1(p_1) \Pr(T \geq s)] + p_1 f_1(p_1) \left[\Pr(T \geq s) - \Pr\left(T + \max\left(p_1 - p_1^E, 0\right) \geq s\right)\right]$. The second term is 0 when $p_1 \leq p_1^E$ and negative otherwise. The derivative of the first term with respect to p_1 is $-f_1(p_1) \Pr\left(T + \max\left(p_1 - p_1^E, 0\right) \geq s\right) - [f_1(p_1) + p_1 f_1'(p_1)] \Pr(T \geq s)$. This is strictly negative provided $-2f_1(p_1) - p_1 f_1'(p_1) < 0$ (this is the second derivative of $p_1 [1 - F_1(p_1)]$). If that holds, the first term is strictly positive for $p_1 \in \left[a_1, p_1^E\right)$ and strictly negative for $p_1 \in \left(p_1^E, p_1^E + s\right]$.[28] (2) $(\star)$ is proportional to

$$D_1 / f_1(p_1) - p_1 \Pr(T \geq s) + p_1 \left[\Pr(T \geq s) - \Pr\left(T + \max\left(p_1 - p_1^E, 0\right) \geq s\right)\right]$$

The second term is 0 when $p_1 \leq p_1^E$ and negative otherwise. Differentiating the first term with respect to p_1 gives $-\Pr\left(T + \max\left(p_1 - p_1^E, 0\right) \geq s\right) - \Pr(T \geq s) - \frac{D_1 f_1'(p_1)}{f_1(p_1)^2}$.

[28] In the case of a corner solution, $p_1^E = a_1$ and therefore we only use the fact that the first term is strictly negative for $p_1 \in \left(p_1^E, p_1^E + s\right]$.

Since $1 - F_1(v)$ is logconcave, this is weakly less than $-2\Pr(T \geq s) + \frac{D_1}{1-F_1(p_1)}$, which is weakly less than $-2\Pr(T \geq s) + 1$, which is negative provided $\Pr(T \geq s) \geq 1/2$. When this holds, profit is strictly increasing in p_1 for $p_1 \in \left[a_1, p_1^E\right)$ and strictly decreasing in p_1 for $p_1 \in \left(p_1^E, p_1^E + s\right]$.[29]

Proof of equation (2.3): Add and subtract $\Pr(T \geq s)\left[1 - F_1\left(p_1^E\right)\right]$ to the lefthand side of (2.2) and rearrange. ■

Claim 1: (2.3) has at most one equilibrium solution: (2.3) is continuous in p_1^E. Any *interior* equilibrium solution must be on the concave part of $p[1 - F_1(p)]$. (From part 1 of the first proof, if p_1^E is on a strictly convex part of $p[1 - F_1(p)]$ then the first order condition is negative for p_1 below but close to p_1^E. This means profit is falling in p_1 for p_1 close to p_1^E so the firm should charge less than p_1^E) On the interval $[a_1, b_1]$, $p[1 - F_1(p)]$ is either all strictly convex, all strictly concave, or there is a $\tilde{p}$ such that it is strictly concave for $p < \tilde{p}$ and strictly convex for $p > \tilde{p}$. (The second derivative is $pf_1(p)\left[-2/p - f'(p)/f(p)\right]$ and the term in square brackets strictly increases in p) When $p[1 - F_1(p)]$ is concave, $1 - F_1\left(p_1^E\right) - p_1^E f_1\left(p_1^E\right)$ is strictly decreasing in p_1^E. So if (**2.3**) has a corner solution, there are no other solutions where $p[1 - F_1(p)]$ is concave. If (**2.3**) has an interior solution, at most one is where $p[1 - F_1(p)]$ is concave. ■

Claim 2: quasiconcavity as $s \to 0$: If p_1^E is interior, it is clear from the first and previous proofs that a necessary and sufficient condition is that p_1^E be on a concave part of $p[1 - F_1(p)]$. (If an increase in n reduces the equilibrium p_1^E, quasiconcavity then

[29] Again, in the case of a corner solution, $p_1^E = a_1$ and therefore we only use the fact that profit is strictly decreasing in p_1 for $p_1 \in \left(p_1^E, p_1^E + s\right]$.

continues to hold) If $p_1^E = a_1$, generically $(\star)$ is strictly negative there. Since $(\star)$ is continuous, it is also strictly negative for $p_1 \in (a_1, a_1 + s]$. ■

Claim 3: the same distribution implies the same price Suppose v_1 and v_2 have the same distribution $F(\cdot)$ but $p_1^E < p_2^E$. This implies:

$$\Pr\left(\sum_{j=1}^n t_j \geq s \geq \sum_{j\neq 1} t_j\right) + \Pr\left(\sum_{j\neq 1} t_j \geq s\right)\left[1 - F\left(p_1^E\right) - p_1^E f\left(p_1^E\right)\right] \leq 0$$

$$\Pr\left(\sum_{j=1}^n t_j \geq s \geq \sum_{j\neq 2} t_j\right) + \Pr\left(\sum_{j\neq 2} t_j \geq s\right)\left[1 - F\left(p_2^E\right) - p_2^E f\left(p_2^E\right)\right] = 0$$

If p_2^E is an equilibrium, then it and p_1^E are on the concave part of $p[1 - F(p)]$. But then $1 - F\left(p_2^E\right) - p_2^E f\left(p_2^E\right)$ is less than $1 - F\left(p_1^E\right) - p_1^E f\left(p_1^E\right)$. Also $\Pr\left(\sum_{j\neq 2} t_j \geq s\right) > \Pr\left(\sum_{j\neq 1} t_j \geq s\right)$, which yields a contradiction. ■

Proof of Proposition 1: for simplicity ignore advertising. For each product j, consider the modified pricing inequality $\frac{1-F_j\left(p_j^E\right)}{p_j^E f_j\left(p_j^E\right)} - \Pr\left(\sum_{k\neq j} t_k \geq s\right) \leq 0$ $(\diamondsuit)$, and call $\tilde{p}_j^E$ the solution when $\Pr\left(\sum_{k\neq j} t_k \geq s\right) = 1/2$. Set $p_j^E = \tilde{p}_j^E \forall j$ and keep adding products until $\Pr\left(\sum_{k\neq j} \max\left(v_k - \tilde{p}_k^E, 0\right) \geq s\right) \geq \frac{1}{2} \forall j$.[30] Suppose this requires m products. The m inequalities of the form $(\diamondsuit)$ give a map $P \to P$ where $P = \times_{k=1}^m \left[p_k^m, \tilde{p}_k^E\right]$. We really care about m inequalities of the form (2.2), whose solutions are (weakly) below those that solve $(\diamond)$. So the former give a continuous map $P \to P$, therefore by Brouwer's fixed point theorem there exists an equilibrium (quasiconcavity is satisfied). This also holds if more products are added. ■

Proof of Corollary 3: Suppose $p_1^E > p_1^m$ and lower it slightly. This increases the number of visitors to the store, which strictly increases demand (and so profits) for

[30] Provided s is finite this happens in finite time. In particular, once it holds for an existing product, it holds for all new products as well.

all other goods. Now think about profits on good 1, and condition on a consumer's T. Show that for any T, profit on good 1 is decreasing in p_1^E. If $T \geq s$, profit on good 1 equals $p_1^E \left[1 - F_1\left(p_1^E\right)\right]$ which is strictly decreasing in p_1^E. If $T = z < s$ and $p_1^E + s - z \geq b_1$, profit on good 1 is zero. If $T = z < s$ but $p_1^E + s - z < b_1$, profit on good 1 is $p_1^E \left[1 - F_1\left(p_1^E + s - z\right)\right]$. The first derivative with respect to p_1^E is $1 - F_1\left(p_1^E + s - z\right) - p_1^E f_1\left(p_1^E + s - z\right)$. This is proportional to $\frac{1-F_1\left(p_1^E+s-z\right)}{f_1\left(p_1^E+s-z\right)} - p_1^E$ which is strictly decreasing in p_1^E. It is negative for $p_1^E = p_1^m$ and therefore negative for all higher p_1^E too. So profit on good 1 is strictly decreasing in p_1^E. ■

Proofs of Propositions 5-7: Let U be the set and cardinality of unadvertised products. p_k^E increases in p_j^E $\forall k \in U, \forall j \neq k$. So if $X = \times_{j=1}^{U} \left[p_j^m, b_j\right]$ the map $X \to X$ is increasing, so by Tarski's fixed point theorem there is a lowest (Pareto dominant) equilibrium. Suppose this equilibrium is non-Diamond, and denoted $\left(q_1^E, ..., q_U^E\right)$. If $q_j^E = p_j^m$ any j then q_j^E is still the equilibrium price when other prices are $\left(q_1^E, .., q_{j-1}^E, q_{j+1}^E, .., q_U^E\right)$ and when new products are added, advertised prices cut, or α parameters increased - so rule this out. *New product* - need to define q_{U+1}^E. If the new product is advertised, q_{U+1}^E is the advertised price; if it's unadvertised, use (2.2) and the vector $\left(q_1^E, ..., q_U^E\right)$ to solve for it. If $Y = \times_{j=1}^{U+1} \left[p_j^m, q_j^E\right]$ the map $Y \to Y$ is increasing so there is a lowest equilibrium, which is lower than before. This is because each of the U old first order conditions is strictly negative when evaluated at $\left(q_1^E, ..., q_U^E, q_{U+1}^E\right)$. *Advertised price* - lowering an advertised price reduces the U first order conditions and so produces a new lowest equilibrium. *Positive demand shocks* - an increase in α_1 just scales demand and

thickness of demand for product 1, but it reduces the first order conditions on all (other) unadvertised products, exactly as in the other two proofs. At all times profit remains quasiconcave in the prices of existing unadvertised goods (see Claim 2).

Relative prices - suppose $\alpha_1 > \alpha_2$ but $p_1^E < p_2^E$. Equilibrium requires $\frac{1-F\left(p_1^E\right)}{p_1^E f\left(p_1^E\right)} + \prod_{j=2}^{n}\left[1-\alpha_j+\alpha_j F\left(p_j^E\right)\right] \leq \frac{1-F\left(p_2^E\right)}{p_2^E f\left(p_2^E\right)} + \prod_{j\neq 2}^{n}\left[1-\alpha_j+\alpha_j F\left(p_j^E\right)\right]$.But $\frac{1-F\left(p_1^E\right)}{p_1^E f\left(p_1^E\right)} > \frac{1-F\left(p_2^E\right)}{p_2^E f\left(p_2^E\right)}$ and $\alpha_2\left[F\left(p_2^E\right)-1\right] > \alpha_1\left[F\left(p_1^E\right)-1\right]$. ■

Proof of Section 2.4.2: Suppose all prices are interior, and focus on good 1. (2.2) slopes down at the equilibrium p_1^E; as when proving Propositions 5-7, we want to show that at the old p_1^E, the first order condition decreases in n and α_j $(j \neq 1)$ and increases in other prices.Divide (2.2) by $\Pr\left(\sum_{j=2}^{n} t_j \geq s\right)$;the ratio $\Pr\left(\sum_{j=1}^{n} t_j \geq s\right)/\Pr\left(\sum_{j=2}^{n} t_j \geq s\right)$ is crucial. C2 ensures it decreases in n. Next expand the top of the ratio by conditioning on each possible v_1. To show the ratio increases in p_2^E say, it is sufficient to show that terms of the form $\Pr\left(\sum_{j=2}^{n} t_j \geq z\right) / \Pr\left(\sum_{j=2}^{n} t_j \geq s\right)$ increase in p_2^E - which they do by condition C1. Now consider an increase in α_2. Take (2.2) and rewrite it as follows (D_z is the lefthand side of (2.2) given shopping cost z and the existence of all n products except product 2)

$$\left(1-\alpha_2+\alpha_2 \Pr\left(v_2 < p_2^E\right)\right) D_s + \alpha_2\left[\Pr\left(v_2 \geq p_2^E + s\right) D_0 + \int_{p_2^E}^{p_2^E+s} f_2\left(z\right) D_{s+p_2^E-z} dz\right]$$

D_0 is negative; if D_s is negative, by C2 so is $C_z \forall z \in (0,s)$. So if p_1^E is interior and (2.2) is zero, $D_s > 0$ and $\Pr\left(v_2 \geq p_2^E + s\right) D_0 + \int_{p_2^E}^{p_2^E+s} f_2\left(z\right) D_{s+p_2^E-z} dz < 0$. Then an increase in α_2 decreases $\Pr\left(v_2 \geq p_2^E + s\right) D_0 + \int_{p_2^E}^{p_2^E+s} f_2\left(z\right) D_{s+p_2^E-z} dz$.[31]

[31] Corner solutions require the proofs to be modified slightly. For example with α_2 comparative statics, if (2.2) is negative, then *potentially* $D_s < 0$. But the (2.2) is a weighted sum of two negative things, so

Now for Proposition 8. When distributions are identical so are unadvertised prices (Claim 3), so there is a lowest equilibrium, and the equilibrium condition is decreasing in the single price. The relevant C2 condition also holds when moving from 2 to 3 products. When $n = 2$ the relevant C1 condition holds so prices are complementary. The α result for $n = 2$ is easiest to prove directly. (Assuming $b_1 - p_1^E > s$), $\Pr(t_1 + t_2 \geq s) / \Pr(t_2 \geq s)$ is

$$\frac{\alpha_1 \Pr(v_1 \geq s)}{\alpha_2 \Pr(v_2 \geq p_2^E + s)} + \alpha_1 \int_0^s h(v_1 = z) \frac{\Pr\left(v_2 \geq p_2^E + s - z\right)}{\Pr(v_2 \geq p_2^E + s)} + 1$$

which clearly decreases in α_2. Using techniques already used, it is also simple to show that very generally (not just for $n = 2$), more popular products are more expensive. ■

stays negative when α_2 is increased.

Chapter 3

Non-Random Search on the Internet

Abstract: Consumer search on the internet is rarely random. Sponsored links appear higher up a webpage and consumers often click them. Firms also bid aggressively for these 'prominent' positions at the top of the page. But why should prominence matter, when visiting an additional website is almost costless? I show that if consumers know what products are available, they usually only search a few firms. Less prominent firms charge higher prices because they only get visited by consumers who value their product highly. This discourages many consumers from searching them, so they earn significantly lower profits than prominent retailers.

3.1 Introduction

In March 2002 Expedia changed the way it displayed flights for United Airlines, after the carrier announced it would no longer pay commissions. United was demoted to the bottom of Expedia search results, and consumers had to make an extra click in order to learn its fares. United's sales plummeted, and less than twenty-four hours later a deal had been struck.[1] Surprisingly the change in layout had a profound effect on demand, even though the cost to consumers of scrolling down the page and clicking United's link was almost zero.

Sponsored ads also demonstrate the importance of visibility on the internet. Google, Yahoo! and Microsoft sell ad space in prominent positions, most notably at the top of their search results. Firms collectively spend around $10 billion on these ads every year, despite the fact that free organic search results are displayed right next to them. Further, amongst sponsored links, those with a higher rank get clicked on more often.[2] As with the United Airlines story, this can at first glance seem surprising. Although consumers often do click sponsored ads[3], they can easily avoid them if they wish. Even when prominent firms are searched, every non-prominent retailer is only one extra click away. So how can a prominent position be so valuable to firms?

I present a simple model in which consumers have heterogeneous preferences and

[1] See for example 'Travel Web Sites Say Airline Deals Don't Affect Searches', The Washington Post April 3, 2002 and 'Orbitz rivals cry foul, claim monopoly in air travel', CNET News.com April 11, 2002.

[2] Ghose and Yang [32].

[3] For example "...over 80 percent of Polish Internet users who recognise sponsored links from other search results, declare that they consciously click on such links" ('Over 80 Percent Web Users Click on Sponsored Links', Polish News Bulletin, March 15, 2006)

search websites in order to gather price and product information. Although making an extra click is essentially costless, prominent firms charge less but earn more profit. Consumers know when they have found their best possible product-price combination, and do not undertake further costly search. Consequently most people will not search very many websites[4], and crucially they will always start by examining firms who are most prominent. Only consumers who have yet to find their best match keep searching - so visitors to a less prominent website have above-average valuations for its product. Less prominent retailers exploit this by charging high prices - which deters consumers from searching them to begin with. Prominent retailers therefore charge less but have much higher demand, and so are more profitable.

Other papers in the literature usually find that when search costs become small, the effect of prominence on price and profits disappears. This is clearly problematic because the two most interesting applications of prominence - internet and supermarkets[5] - would typically be expected to have only small search costs. The three most closely related papers are those by Armstrong *et al* [4], Haan and Moraga-Gonzalez [34] and Zhou [87]. In Armstrong *et al* one firm is prominent and always searched first; in Haan and Moraga-Gonzalez two firms compete for the right to be prominent by sending out adverts. A prominent firm charges less, and the intuition is the same as in the current paper.[6]

[4]This also seems true in practice. For example one in four Polish Internet users never visits more than three websites before making a purchase (see Polish News Bulletin). Koulayev [47] says that only one third of searchers view more than one page of hotel results.

[5]An extension of the current model shows that products left in prominent end-of-aisle supermarket displays, should have lower prices but earn more profits.

[6]Two other papers by Arbatskaya [2] and Wilson [85] argue that prominent firms charge *higher* prices because they have a monopoly over consumers with high search costs. However this consideration is

However crucially, consumers have no idea how much a product might be worth until they have investigated it. When the search cost becomes small, the gains from visiting an additional retailer outweigh the costs, so (almost) every consumer visits every firm. Search order is then irrelevant, and prominence has no effect on market outcomes. My paper makes the more realistic assumption that often consumers have some knowledge about what products are offered in a market - and therefore do not need to search every firm, even when search is very easy. It therefore explains why prominence can matter even in near frictionless environments like the internet. Zhou [87] explores a very similar idea in a model with two firms on a Hotelling line. Consumers know their own locations but not those of the firms, and choose probabilistically which store to visit first. A difficulty is that generally one of the firms is cheaper than the other, and so one might expect (especially on the internet) that rational consumers would all begin by searching the cheapest retailer. But in this case the Diamond Paradox causes the model to break down. The set-up in my paper avoids this difficulty.

Finally, Taylor [70] and White [84] study the interactions between sponsored and organic search results. They argue that search engines compete for visitors by improving the quality/relevancy of their organic links, which then cuts into advertising revenue. In Taylor links do not set prices, but have varying levels of relevancy. Organic (effectively less prominent) links are more relevant and are clicked first by consumers. In the current paper there is a single search engine and so this quality dimension is not present. However

probably less important on the internet, where most consumers could be expected to have low search costs.

I find that when price competition is introduced, prominent firms charge lower prices - and hence could be viewed as having higher quality. I also demonstrate that improvements in search algorithms lower consumer surplus; by better matching consumers with desirable products, they also cause retailers to increase their prices.

3.2 Assumptions

There are n differentiated products, and a unit mass of potential consumers. Consumers have unit demand and wish to buy at most one product. Each consumer cares only about two randomly selected products. If a consumer likes products i and j, he values these two products at v_i and v_j respectively. It is convenient to work with the difference of these two valuations $d = v_i - v_j$. d is distributed on an interval $[-D, D]$ with a distribution function $H(d)$. The density $h(d)$ is symmetric, continuously differentiable, and logconcave. Each consumer knows which two products he likes, as well as the specific valuations v_i and v_j he attaches to them. Consequently each consumer also knows his own d value. As is standard, product valuations are assumed to be sufficiently high that each consumer always makes a purchase.[7]

There are n firms, each of which produces one of the differentiated goods at zero cost. Consumers do not know which firm sells which product, and nor do they know how much each firm charges. This information must be acquired by consumers through costly sequential search. Each firm has a website giving details about its product and price.

[7] This generalises Hotelling's linear city model and Chen and Riordan [18].

Visiting a particular firm's website and learning this information costs the consumer s. To contrast with other papers, I assume that $s \to 0$ but remains positive. Consumers have costless recall and can costlessly buy from any firm whose website they have already searched.

A good example might be flights - consumers have preferences over price and time. They know roughly when flights might take off, but not which airline occupies which time slot. In reality consumers always have *some* small uncertainty about what products are available, and their search cost is small but certainly not tending to zero. Nevertheless this situation of small uncertainty and small search costs is hopefully well-approximated by perfect knowledge about products and $s \to 0$.

Finally there is a platform which provides links to each of the n websites. These links can be arranged in one of two ways: **with prominence** and **without prominence**.

Without prominence, all firms are treated equally. Each consumer receives a randomly generated list of n firms. A firm never learns its position in any list, and therefore charges each consumer the same price.

With prominence, some firms are treated more favourably. There are two groups - a prominent group g_1 and a non-prominent group g_2. They contain n_1 and $n_2 = n - n_1$ firms respectively. A firm always appears in the same group, and knows which group it is. Within a group, firms are treated equally and randomly assigned a list position. When setting their price, firms can only condition on whether they will appear in g_1 or g_2. At present 'prominent' and 'non-prominent' are just labels. However I demonstrate

that the platform can easily ensure that g_1 firms are always searched first by consumers. In this sense they are prominent.

3.3 Equilibrium Pricing Behaviour

3.3.1 Without Prominence

I look for a symmetric equilibrium in which each firm charges a price p^*. Consumers have passive beliefs - meaning that if they visit a website and the firm charges something different from p^*, consumers still expect other (as yet unsearched) firms to charge p^*.

I first demonstrate that optimal consumer search takes the form of a simple cutoff rule. Each consumer who visits the platform is presented with a randomly-generated list. Denote by i and j the two products that the consumer likes, and suppose that product i is higher in the list. Without loss of generality suppose consumers search from the top of the list downwards. Assume also that product i is located at the l^{th} website and is priced at p. Let p_j denote the actual price of product j (whilst p^* is its expected price). Then the optimal search rule is:

Lemma 10 ***Optimal 'without prominence' search rule***

A *Visit websites until product i is found. Then*

B *i)* *If* $v_i - p \geq v_j - p^* - s\frac{n-l+1}{2}$ *buy i immediately*

ii) *If* $v_i - p < v_j - p^* - s\frac{n-l+1}{2}$ *search until finding j*

Then buy j iff $v_i - p < v_j - p_j$, otherwise buy i

A consumer should always visit websites until finding one of his two favourite products (part A), because by assumption search costs are minimal and valuations are large relative to the expected price p^*. Once product i has been found, the consumer either buys good i immediately, or keeps searching through the list until finding good j (part B). Intuitively, suppose that the consumer, *behaving optimally*, is just indifferent between consuming good i immediately or visiting the $l+1^{th}$ website. If the $l+1^{th}$ website is clicked on, either firm j is found immediately (in which case no further search occurs), or the consumer learns that it is in one of the remaining websites. In the latter case, search becomes more favourable than it was previously, because the $l+1^{th}$ website has been 'eliminated'. So if the consumer was previously indifferent about clicking on the $l+1^{th}$ website, once $l+1$ is eliminated, he strictly prefers to keep searching until finding product j. Consequently the optimal search decision compares the payoff from consuming product i immediately $(v_i - p)$, and the expected payoff of searching every other link until finding good j $(v_j - p^* - s\frac{n-l+1}{2})$.[8]

The decision about whether to buy product i immediately or search on for product j is made based upon the *actual* price of i and the *expected* price of j. The consumer is called a *shopper* for product i and a *Diamond consumer* for product j.

Recall that within the literature, a *shopper* is somebody who has a zero search cost. Suppose $s = 0$ and i's rival charges p^*. Then demand for product i is equal to $1 -$

[8]The cost of clicking links until finding product j is given by $s\frac{n-l+1}{2} = s\sum_{z=1}^{n-l}\frac{z}{n-l}$.
This differs from Weitzman [81], where the optimal search rule is myopic. The current search rule is non-stationary because previous searches reveal information about the benefits of future search. This complicates the analysis, and therefore I focus only on small search costs.

$H(p-p^*)$, and the slope of demand is $-h(p-p^*)$. Now suppose instead that $s>0$ and i's rival charges p^*. The consumer either buys product i immediately or he never buys it. Therefore demand for product i equals $1-H\left(p-p^*-s\frac{n-l+1}{2}\right)$, and its slope is $-h\left(p-p^*-s\frac{n-l+1}{2}\right)$. Demand essentially behaves the same when $s=0$ and when $s\to 0$. The reason is that in both cases, the decision about whether or not to buy product i is made based upon its actual price. This is the reason for the 'shopper' terminology.

The *Diamond consumer* terminology is similar but not identical to that used in the previous Chapter. Take the case $s>0$. In equilibrium firm i charges p^* so a consumer only searches for product j if $v_i-v_j<-s\frac{n-l+1}{2}$ and then only buys j if $v_i-v_j<p^*-p_j$. So demand for product j is $\min\left[H\left(-s\frac{n-l+1}{2}\right),H\left(p^*-p_j\right)\right]$. Notice that this demand is *perfectly inelastic* for small increases in actual price p_j above expected price p^* - just as in Diamond's [20] paradox. Intuitively consumers with $v_i-v_j\geq -s\frac{n-l+1}{2}$ never search for product j so they never learn p_j and cannot respond to it. Consumers with $v_i-v_j<-s\frac{n-l+1}{2}$ do search for product j and they do learn p_j. However they search for it precisely because it gives them a strictly better deal than product i does. This means that if firm j increases price a little above the expected level p^*, all these consumers are still willing to buy there. Consequently demand behaves very differently depending upon whether $s=0$ or $s\to 0$. The reason is that when $s\to 0$, the equilibrium decision about buying j or not is based upon its expected price, not its actual price.

Lemma 11 *Without prominence, the equilibrium price is* $p^*=\frac{1}{h(0)}$

In equilibrium every firm charges $p^* = \frac{1}{h(0)}$, and every consumer expects this. Consider a firm which deviated by charging a price slightly higher than p^*. Most consumers would pay the higher price, so the firm would gain by an amount equal to $\frac{1}{2}$, which is just the average demand earned on each consumer interested in its product. Some shoppers for the product would stop buying, but any Diamond consumer who intended to make a purchase would still do so. Since lists are generated randomly, half the time somebody is a shopper for the firm, and half the time they are a Diamond consumer. So the mass of consumers who would substitute away is $\frac{1}{2}h(0)$. The firm's first order condition is then $\frac{1}{2} - \frac{1}{2}p^*h(0) = 0$, which obviously gives $p^* = \frac{1}{h(0)}$. Notice that p^* is precisely double the price that firms charge in a standard Hotelling model. Intuitively firms get the same demand in both cases, but when s is positive, demand is only half as responsive to small price changes.

To summarise, when we go from $s = 0$ to $s \to 0$, equilibrium price doubles. Therefore imposing arbitrarily small costs of acquiring extra information has large effects on price, even in a symmetric setting. The reason is that some search decisions are taken based upon expected rather than actual prices, which makes demand curves much less elastic. I now show how small information frictions interact with prominence.

3.3.2 With Prominence

Each firm is now placed into either a prominent group g_1 or a non-prominent group g_2. Recall that within a group, each firm is treated equally and list position is determined

randomly. The groups contain n_1 and $n_2 = n - n_1 \geq 2$ websites. Consumers have passive beliefs, and expect firms in the two groups to charge prices p_1^E and p_2^E respectively. I again consider the limiting equilibrium as $s \to 0$.

I first look for an equilibrium in which expected prices satisfy $p_1^E < p_2^E$. Consumers should only visit a g_2 retailer once every g_1 firm has been visited and found unsatisfactory. Place links into a list, with g_1 firms occupying the top n_1 positions. Then without loss of generality assume that consumers search the list from top to bottom. Using the same notation as in the without prominence case, denote by i and j the two products that the consumer likes. Suppose that product i is higher in the list, is located at the l^{th} website, and is priced at p. Let p_j denote the actual price of product j. Recall that consumers have passive expectations, that p_1^E is the price expected to be found at any g_1-retailer, and p_2^E is the price expected to be found at any g_2-retailer. Then for s sufficiently small, the following holds:

Lemma 12 ***Optimal 'with prominence' search rule***

A *Visit websites until product i is found. Then*

B *If i is in g_1 at position $l = n_1$, then go to Part B2ii)*

If i is in g_1 at position $l < n_1$, then:

1 *Buy i immediately if $v_i - p \geq v_j - p_1^E - s\left(\frac{n_1 - l + 1}{2} + n_2\right)$*

2 *Search for j if $v_i - p < v_j - p_1^E - s\left(\frac{n_1 - l + 1}{2} + n_2\right)$*

i) *If j is found in g_1, buy j if $v_i - p < v_j - p_j$*

ii) *If j is not found in g_1, then*

- *Buy i immediately if $v_i - p \geq v_j - p_2^E - s\frac{n_2+1}{2}$*
- *Search in g_2 for j if $v_i - p < v_j - p_2^E - s\frac{n_2+1}{2}$*

Once j is found, buy j if $v_i - p < v_j - p_j$

C *If i is in g_2, then:*

1 *Buy i immediately if $v_i - p \geq v_j - p_2^E - s\frac{n-l+1}{2}$*

2 *Search for j if $v_i - p < v_j - p_2^E - s\frac{n-l+1}{2}$*

Once j is found, buy j if $v_i - p < v_j - p_j$

It is again optimal to search at least until finding product i (Part A). If product i is in the second group, then so is product j. Consequently the search rule is the same as that described earlier in the case without prominence (Part C). If instead product i is in the first group, search behaviour is a little richer (Part B). If product i is a reasonable deal, the consumer buys it immediately without further search (Part B1). Otherwise the consumer searches the rest of g_1 hoping to find product j. Suppose that g_1 has been searched but product j has not been found. Depending upon how bad the deal

on product i is, the consumer either buys good i immediately (with no further search) or searches through g_2 until finding product j (Part B2ii). When $s \longrightarrow 0$ the search rule takes a particularly simple form: click the next link provided that if product j is found there, it is expected to be strictly superior to i.

Consumers search prominent firms first because they expect them to be cheaper. A prominent firm is therefore more likely to be seen first by consumers who are interested in its product. These consumers are more likely to know the firm's actual price when they are deciding whether or not to search on. Consequently prominent firms are better able to increase their demand by cutting price. In contrast a non-prominent firm is less likely to be seen first by consumers who are interested in its product. These consumers are more likely to make their search decision based only upon an expectation of the firm's price. Hence their demand is not very sensitive to price changes. It then follows:

Proposition 13 *There is a unique equilibrium with $p_1^E < p_2^E$. Prices satisfy*

$$\begin{aligned} p_1^E &= \frac{n_1 - 1 + 2n_2\left[1 - H\left(p_1^E - p_2^E\right)\right]}{\left[n_1 - 1\right]h\left(0\right) + 2n_2 h\left(p_1^E - p_2^E\right)} \\ p_2^E &= \frac{2n_1 H\left(p_1^E - p_2^E\right) + n_2 - 1}{\left[n_2 - 1\right]h\left(0\right)} \end{aligned}$$

Compared to their non-prominent competitors, prominent firms have more shoppers and less Diamond consumers. They face a more elastic demand curve and hence charge lower prices. The expressions in Proposition 13 are only valid when $n_2 \geq 2$. Mathematically this is because if $n_2 = 1$, the expression for p_2^E requires us to divide by 0.

Intuitively if g_2 contains a single firm, then everybody is a Diamond consumer for its product. Anybody searching the firm signals that they expect it to give strictly more utility than is available at their other alternative. This means the firm can always increase its profit by charging more than expected. Equilibrium would then require p_2^E to be so large that no consumer ever finds it optimal to visit the last retailer.

Proposition 13 differs from Armstrong *et al* [4] because in their paper prominence has no effect when $s \to 0$. In their model, product valuations are drawn independently from a distribution. Prior to searching a firm, the consumer has no information about what its product is worth. The potential benefits of searching an additional firm are therefore large, whilst the costs become trivial when $s \to 0$. Therefore (almost) every consumer searches every firm, and learns all prices. Search order becomes unimportant, and all firms charge the same price. In reality consumers probably know quite a lot about the products that are available in the market. My model attempts to capture this by assuming that everybody knows product valuations, but does not know which firm sells which product. Consequently consumers know when they have found their best product, and do not need to engage in further costly search. Almost no consumer needs to search the entire list, so prominence matters even when $s \to 0$.

Proposition 13 also differs from Armstrong *et al* [4] in terms of what happens when n becomes 'large'. In particular, suppose $n_1 = 1$ and n_2 is large. In their model all firms charge (almost) the same price regardless of s. But in my model g_2 firms essentially play only against themselves and therefore charge p^*, whilst the prominent retailer sets

a best response to p^* (and so still charges significantly less, since p^* is double the usual Hotelling price).

Remark 14 *Compared to the situation where no firms are prominent, and everyone charges* $p^* = \frac{1}{h(0)}$

- *Non-prominent firms charge more*

- *Prominent firms may charge more or less*

Without prominence, each firm has on average an identical number of shoppers and Diamond consumers. With prominence, the firms in g_2 have relatively more Diamond consumers than shoppers. If a firm charges a slightly higher price than expected, it gains on Diamond consumers but loses on shoppers. The gains are independent of the price level, but the losses are not. In particular, when price is higher, each marginal consumer (i.e. shopper) is more valuable. To ensure the gains and losses balance, the equilibrium p_2^E must exceed $\frac{1}{h(0)}$. Now think about the price charged by prominent firms. On the one hand g_1 retailers have more shoppers relative to Diamond consumers, face a more elastic demand curve, and therefore want to charge less than $\frac{1}{h(0)}$. On the other hand, their rivals in g_2 now charge more, which due to strategic complementarity encourages firms in g_1 to charge more as well.[9] Either effect can dominate.

[9]It is interesting to rewrite the expression for p_1^E. When $n_1 = 1$, $p_1^E = \frac{1-H\left(p_1^E - p_2^E\right)}{h\left(p_1^E - p_2^E\right)}$ so the g_1 retailer sets a best response to a rival who charges p_2^E. When $n_1 > 1$, $p_1^E = \alpha \frac{1}{h(0)} + (1-\alpha)\frac{1-H\left(p_1^E - p_2^E\right)}{h\left(p_1^E - p_2^E\right)}$ with $\alpha \in [0,1)$; p_1^E is an average of $\frac{1}{h(0)}$ and a best response.

3.3.3 Welfare and Profit Comparison

I begin with two numerical examples in which $n_1 = 1$ and $n_2 = 5$ The accompanying table shows how key market outcomes change when we move from the 'without prominence' case (in which all six firms are searched randomly) to the 'with prominence' case (where $n_1 = 1$ and $n_2 = 5$).

	Percentage Effect of Prominence	
	$d \sim U[-D, D]$	$d \sim N[0, 1]$
Prominent firm's price	-20	-11.31
Non-prominent firm's price	$+10$	$+13.39$
Prominent firm's demand	$+60$	$+46.43$
Non-prominent firm's demand	-12	-9.29
Prominent firm's profit	$+28$	$+29.86$
Non-prominent firm's profit	-3.2	$+2.86$
Total industry profit	$+2$	$+7.36$
Total surplus	$\downarrow$	$\downarrow$
Expected number of clicks	-8.57	-6.63

What is immediately striking is just how big many of the effects are. For example the prominent firm may reduce its price by as much as 20%, and earn 30% more profit compared to the situation when it was searched randomly. Moreover prominence causes large distortions in prices, which result in a high amount of substitution towards the g_1 retailer (and a reduction in total clicks). It is important to remember that only $\frac{1}{3}$ of

consumers are actually interested in the prominent firm's product. Of those, 80% (in the Uniform case) and 73.2% (in the Normal case) end up buying its product. In line with the discussion in the previous section, in one example non-prominent firms earn more whilst in the other they earn less. It is also simple to build Normal examples in which p_1^E increases. However all other variables move in the same direction regardless of the particular distribution - something which I show in the rest of this section.

Lemma 15 *Prominent firms earn more profit than non-prominent firms*

A prominent position is valuable and firms are willing to bid aggressively to acquire one. Even though s is almost zero, consumers have lots of match information and therefore do not typically search very far. As was shown in the previous section, this creates a significant distortion in prices. One implication of this distortion is that, ceteris paribus, g_1 retailers face a larger demand than g_2 retailers.[10] It then follows that if a g_1 retailer deviates from equilibrium and charges a price p_2^E, it earns strictly more profit than a g_2 retailer does. By revealed preference g_1 retailers actually charge less than p_2^E and therefore do better still. The example in the earlier Table illustrates that the benefits of being prominent can be extremely large.

Proposition 16 *Industry profit is higher when some firms are prominent*

Prominence distorts prices in such a way that firms can collectively extract higher payments from consumers. The result is obvious when $p_1^E \geq \frac{1}{h(0)}$, so suppose instead

[10] In equilibrium a g_1 retailer faces $n_1 - 1$ rivals who charge p_1^E and n_2 rivals who charge p_2^E; a g_2 retailer faces n_1 rivals who charge p_1^E and $n_2 - 1$ rivals who charge p_2^E.

that prominence makes g_1 firms cheaper. The proof proceeds by way of the following experiment. Fix equilibrium prices, and make each consumer buy from the firm for which they are a shopper. A retailer with more shoppers has higher demand, but also charges proportionately lower prices (shoppers enter into the denominators of the price expressions in Proposition 13). It is then easy to show that under this experiment, industry profits weakly exceed $\frac{1}{h(0)}$. The profits under the experiment are themselves a strict lower bound on actual industry profits.

Retailers in g_1 earn more profit than they would if consumers searched randomly. This is because the prominent firms get a larger share of a larger industry profit. The effect on non-prominent firms is ambiguous, because they get a smaller share of a larger pie. When d is uniform, non-prominent firms always earn a profit below $\frac{1}{nh(0)}$ and therefore are made worse off. However the $N(0,1)$ example in the earlier Table shows that sometimes even less prominent firms can benefit.

Society is made worse off when some firms are made prominent. Total welfare is maximised (in the limit) when consumers search randomly. This is because all firms charge the same price, so each consumer buys the product which gives him the highest gross surplus. When some firms are made prominent $p_1^E < p_2^E$, so 'too many' people buy from prominent retailers and therefore total welfare is lower. Consumer welfare is also lower - consumers buy the 'wrong' products, and end up paying more for them. But given that everybody else follows the search order, any individual consumer should do the same.

I end this section by briefly comparing the main comparative statics results with those in Armstrong *et al* [4]. As explained earlier, in their paper prominence has no effect when s becomes very small, and therefore I compare my results (as $s \to 0$) with the results they derive for larger values of s. In their model, prominent firms face more elastic demand curves and therefore also charge lower prices, although their intuition is given in terms of 'fresh' and 'return' demand rather than shoppers and Diamond consumers.[11] They also find that that non-prominent firms charge more than they would in the symmetric situation; unlike in my model, they also show that prominent firms always charge less. In both models prominence lowers total and consumer welfare, raises industry profit, and has an ambiguous effect on the profits of non-prominent retailers.

3.3.4 Equilibrium Selection

Up until now, we have exogenously assumed that consumers expect prices to satisfy $p_1^E < p_2^E$. If we do not make this assumption, there are actually three possible equilibrium outcomes. Firstly - as shown in Proposition 13 - there is one equilibrium where consumers expect $p_1^E < p_2^E$. Secondly, there is one equilibrium where consumers expect $p_1^E > p_2^E$. Search behaviour is qualitatively the same as in Lemma 12, except that consumers begin by clicking links in the bottom group. The equilibrium values of p_1^E and p_2^E satisfy two equations that are similar to those in Proposition 13. The third and final equilibrium outcome is that consumers expect $p_1^E = p_2^E = \frac{1}{h(0)}$. Consumers are willing to search

[11] Note that this is more than just a difference in terminology: a shopper and fresh consumer are not the same, and nor are a Diamond and return consumer.

randomly, and therefore Lemma 11 is valid.[12] Given the way the model has been constructed, the terms 'prominent' and 'non-prominent' are merely labels, and have no necessary bearing on which equilibrium is played. If the platform wishes to select the equilibrium with $p_1^E < p_2^E$, there are three possible ways to do this:

1. Enforce a search order

2. Use context to coordinate consumers

3. Sell the rights to prominence

The platform may force consumers to search the firms in g_1 first. Precisely, a consumer can only visit a g_2 retailer once he has first looked at every g_1 firm.[13] This game has a unique equilibrium - the same one as identified in Proposition 13. In particular there cannot be an equilibrium with $p_1^E \geq p_2^E$, because firms in the top group would still be searched first, and would therefore continue to have more shoppers. Their demand would be more elastic and they would have incentives to charge lower prices than retailers in the bottom group.[14]

Context may coordinate consumers on one equilibrium, and Google is a good example of how this can be done. Google sponsored ads are placed in a yellow box and therefore

[12] However there are many ways consumers can search, such that firms charge $\frac{1}{h(0)}$. In particular, all that is required is that on average each firm receives the same number of shoppers and Diamond consumers.

[13] Armstrong *et al* [4] give the example of travel agents, who choose the order in which they present options to you.

[14] Suppose for example that consumers expected $p_1^E > p_2^E$. Their optimal search rule would be: buy i immediately, or search the entire list until finding j. As $s \to 0$, a consumer would search provided $v_j - p_2^E > v_i - p_i$. In particular if j were found in g_1 at price p_1^E, it would be unclear if the consumer preferred i or j. Consequently *everybody* would be a shopper for firms in g_1, which would make their demand very elastic.

made to stand out from organic results. The most expensive sponsored ads are placed at the top of the page, which is typically where consumers look first. Furthermore organic search results are sometimes separated from the top sponsored ads by way of a large white space. All of these actions help sponsored ads to stand out, and potentially coordinate consumers by encouraging them to search these retailers first.

Perhaps the most effective way of coordinating consumers is to sell off the right to be prominent. Suppose that in the $p_1^E < p_2^E$ equilibrium, retailers in the two groups make profits Π_{g_1} and Π_{g_2} respectively. Suppose further that the platform sells a slot in g_1 for $\Pi_{g_1} - \epsilon$ and a slot in g_2 for $\Pi_{g_2} - \epsilon$, and that this is public information. Consumers should realise that firms only pay extra to be prominent because they expect the $p_1^E < p_2^E$ equilibrium. It is then rational for consumers to begin their search by looking at g_1 retailers, in which case firms' expectations are fulfilled. More generally, if consumers realise that retailers pay extra to be prominent, this should signal to consumers that they are expected to play the equilibrium with prominence.

In summary, the platform has several ways in which it can coordinate play on the equilibrium in which g_1 retailers are prominent. When such an equilibrium is played, consumers are rational in searching prominent retailers first.[15]

[15] It is important that a) there are only two groups and b). each consumer gets a randomly generated list of firms within a particular group. This implies there are just the three equilibrium outcomes. If instead consumers were presented with the same list, the set of equilibria would be much richer, and need not be confined to one or two groups. Another implication is that if the platform enforces a search order, the search order *must* treat firms within a group in a symmetric way.

3.4 Increasing the Degree of Prominence

We have seen that knowledge of products coupled with small information frictions leads to large price distortions. In this section I move away from the internet interpretation, and investigate under what circumstances prominent firms could be more expensive. Consumers still pay a small cost $s \to 0$ to search a firm and learn its price and product. However before *any* g_2 retailer can be searched, the consumer must pay a one-off cost $\tilde{s} \in [0, D)$. On Google $\tilde{s} \approx 0$; travel agents on the other hand often promote certain holidays, and may be reluctant to give details on other packages - in which case $\tilde{s}$ could be non-trivial. This section demonstrates that when $\tilde{s}$ is large enough, prominent firms will be more expensive, but that the required $\tilde{s}$ is usually very large - so in most interesting applications, prominent firms should be expected to be cheaper. $\tilde{s}$ may also be viewed as a tractable way of introducing higher search costs.

I again look for equilibria in which consumers expect prices to satisfy $p_1^E < p_2^E + \tilde{s}$. Given their expectation, consumers always begin their search in the prominent group, and only visit the g_2 retailers when there are no more g_1 firms left to search. The optimal search rule is identical to that in Lemma 12 except that when deciding whether or not to begin searching g_2, the consumer uses an effective price $p_2^E + \tilde{s}$ rather than just p_2^E. It is then simple to show that there exists a unique equilibrium, again almost the same as

that identified in Proposition 13, with prices satisfying

$$\begin{aligned} p_1^E &= \frac{n_1 - 1 + 2n_2\left[1 - H\left(p_1^E - p_2^E - \tilde{s}\right)\right]}{\left[n_1 - 1\right]h\left(0\right) + 2n_2 h\left(p_1^E - p_2^E - \tilde{s}\right)} \\ p_2^E &= \frac{2n_1 H\left(p_1^E - p_2^E - \tilde{s}\right) + n_2 - 1}{\left[n_2 - 1\right]h\left(0\right)} \end{aligned}$$

Although prominent firms now have some monopoly power courtesy of the $\tilde{s}$ cost, they still have more shoppers and less Diamond consumers. Consequently they are still cheaper once $\tilde{s}$ is accounted for. Using the same method as before, industry profit is still higher compared to the 'without prominence' case, regardless of the size of $\tilde{s}$. Prominence also still reduces total and consumer welfare - there is now a third channel, namely that whenever $\tilde{s}$ is paid, it is a deadweight social loss. Of most interest, therefore, are comparative statics in $\tilde{s}$.

Lemma 17 *When $\tilde{s}$ increases:*

- *p_1^E increases and p_2^E decreases*
- *Firms in g_1 earn more profit, whilst firms in g_2 earn less*

A larger $\tilde{s}$ acts to discourage some consumers from searching the bottom group. Other things equal, this increases demand for g_1 retailers and decreases demand for g_2 retailers. Therefore firms in the top group increase their price a little to take advantage of this extra monopoly power; firms in the bottom group reduce their price since they now have fewer Diamond consumers searching for them from the prominent group. Equilibrium

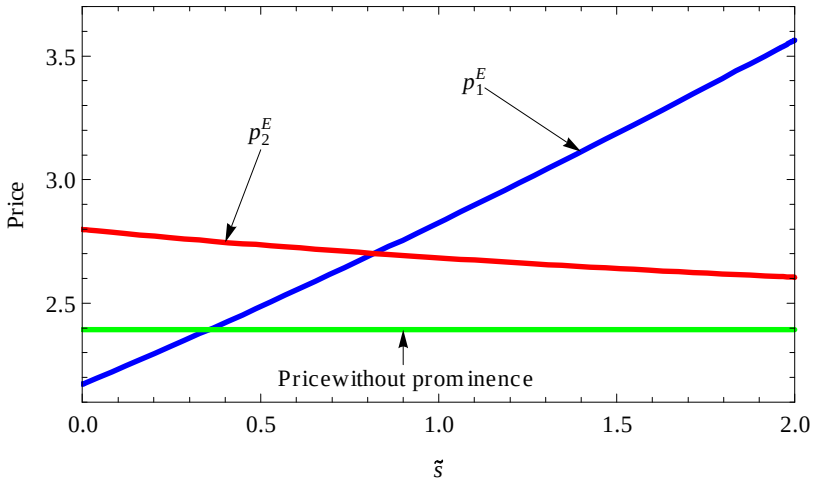


Figure 3-1: How prices are affected by changes in $\tilde{s}$

prices therefore reflect two opposing forces - an information effect, and a monopoly effect. Prominent firms infer less about a consumer's valuation for their product, since they have more shoppers and less Diamond consumers. This makes them want to charge less. However prominent firms also have some monopoly power since searching g_2 is made exogenously less attractive. This encourages the prominent retailers to charge higher prices. When d is uniform, the information effect *always dominates*, and $p_1^E < p_2^E$ for any $\tilde{s} < D$. However for non-uniform distributions, one can always find an $\tilde{s} < D$ such that $p_1^E = p_2^E$, although usually $\tilde{s}$ needs to be quite large. Figure 3-1 has an example in which $n_1 = 1$, $n_2 = 4$ and valuations are standard normal but truncated so they fit on the interval $[-2, 2]$. When $\tilde{s} = 0$ prominent firms charge less than in the symmetric case, but once $\tilde{s}$ reaches around $\frac{4}{5}$ prominent firms begin to charge more. However by this stage only 20% of consumers who liked one prominent and one non-prominent product would have a strong enough valuation to want to search into the bottom group.

Even accounting for changes in prices, consumers substitute towards prominent retailers when $\tilde{s}$ increases. These prominent firms therefore charge a higher price *and* enjoy a higher demand - and therefore unambiguously benefit when $\tilde{s}$ increases. For precisely the opposite reasons, non-prominent firms are made unambiguously worse off when $\tilde{s}$ increases. Overall industry profit behaves as follows:

Proposition 18 *The effect of $\tilde{s}$ on industry profit is proportional to*

$$p_1^E - p_2^E + \frac{1}{2} p_1^E p_1^E \frac{h'\left(p_1^E - p_2^E - \tilde{s}\right)}{h\left(p_1^E - p_2^E - \tilde{s}\right)}$$

A change in $\tilde{s}$ affects industry profits through three channels. Firstly, some marginal consumers substitute towards the prominent retailers, so the change in their expenditure is proportional to $p_1^E - p_2^E$. In addition, each retailer's price is equal to its demand divided by its thickness of demand. The second effect - also proportional to $p_1^E - p_2^E$ - is that $\tilde{s}$ reallocates demand and therefore changes prices. The third effect occurs because a g_1 retailer's thickness of demand has a term $h\left(p_1^E - p_2^E - \tilde{s}\right)$, so reallocations in demand also change their prices through this channel. When $h(d)$ is uniform, industry profit is monotonically decreasing in $\tilde{s}$. Indeed industry profits are the same when (i) all firms are treated equally and (ii) there is prominence but $\tilde{s} \to D$ (although the distribution of profits is very unequal in the second case). When $h(d)$ is non-uniform, industry profit is U-shaped in $\tilde{s}$[16] and maximised when $\tilde{s}$ is either zero else equal to D.

[16] $p_1^E - p_2^E$ is increasing in $\tilde{s}$; $h'\left(p_1^E - p_2^E - \tilde{s}\right)$ is positive and by logconcavity $\frac{h'(x)}{h(x)}$ is decreasing in x,

Lemma 19 *Welfare is typically decreasing in* $\tilde{s}$

Consumer surplus is definitely decreasing in $\tilde{s}$, and usually total surplus is as well. Consumers pay p_1^E or $p_2^E + \tilde{s}$ depending upon whether they buy from a g_1 or a g_2 retailer; both these (effective) prices are increasing in $\tilde{s}$. Most consumers do not change their purchase decisions, and so pay more money. Some marginal consumers switch from a g_2 product to a g_1 product, but they must have been indifferent about what to buy, and so do not gain by switching. Therefore consumer surplus decreases in $\tilde{s}$. Total surplus is affected by $\tilde{s}$ in two ways. The first - negative effect - is that most people who were buying a g_2 product continue to do so, therefore total $\tilde{s}$ costs increase. The second - ambiguous - effect is that some consumers switch from a g_2 retailer to a g_1 retailer. This switching adds an amount $v_1 + \tilde{s} - v_2$ to total surplus; since these consumers were indifferent and therefore had $v_1 - p_1^E = v_2 - p_2^E - \tilde{s}$, the gain to society is actually $p_1^E - p_2^E$. Intuitively if $p_1^E < p_2^E$ then too many consumers are buying a prominent product. In this case allocating more consumers to a prominent retailer makes a bad situation worse. When $h(d)$ is uniform we know that $p_1^E < p_2^E$ so increasing $\tilde{s}$ is unambiguously bad for welfare. More generally it is definitely bad when $\tilde{s}$ is small, and probably bad even when it gets large.

---

so since $p_1^E - p_2^E - \tilde{s}$ is decreasing in $\tilde{s}$, the term $\frac{1}{2} p_1^E p_1^E \frac{h'\left(p_1^E - p_2^E - \tilde{s}\right)}{h\left(p_1^E - p_2^E - \tilde{s}\right)}$ also increases in $\tilde{s}$. So either (a) profit is always increasing in $\tilde{s}$ or (b) profit falls at first but then rises in $\tilde{s}$.

3.5 Discussion

3.5.1 The Importance of What Consumers Know

The effect of small search costs depends critically on how much consumers know about the set of available products. In Armstrong *et al* [4] consumers have no information about how much they value unsearched products. When s becomes small, an additional search is (almost) always worthwhile and so most consumers view everything before making a purchase. Firms therefore price, essentially, as if there were no search cost. An alternative view is that consumers have a good idea of what products are available, and know when they have found the best deal (for them, given their preferences). Even in a symmetric setting, small search costs drastically relax price competition (Section 3.3.1), whilst prominence leads to large distortions in prices and profits (Section 3.3.2).

There are clear applications to the burgeoning empirical literature, which estimates online search to be very expensive. For example Koulayev [47] finds that the median consumer pays a search cost of \$37.58 when processing a results page with fifteen hotels on it. In general consumers search little whilst price variation is significant. Large search costs help rationalise these facts, but seem inconsistent with the near-frictionless world of e-commerce. The current paper suggests that by modelling consumer information, one could rationalise these facts and still estimate a much smaller search cost.

A simple extension of the model also demonstrates that improvements in search algorithms harm consumers. Suppose that $1-x$ of consumers behave as assumed in Section

3.2, whilst the remainder know which firm they like best and therefore go straight to its website. Then prices satisfy

$$\begin{aligned} p_1^E &= \frac{n_1 - 1 + 2n_2\left[1 - H\left(p_1^E - p_2^E\right)\right]}{(1-x)\left[(n_1-1)\,h\,(0) + 2n_2 h\,(p_1^E - p_2^E)\right]} \\ p_2^E &= \frac{2n_1 H\left(p_1^E - p_2^E\right) + n_2 - 1}{(1-x)\,(n_2-1)\,h\,(0)} \end{aligned}$$

and are increasing in x.[17] One could also interpret x as a measure of how successful a search engine is at directing consumers to their preferred product. When the matching technology is better, more of a website's visitors definitely like its product - each retailer has more Diamond consumers and therefore charges more. This assumes however that product valuations are large; more generally some consumers would respond to price increases by exiting the market. Indeed as $x \to 1$, the Diamond Paradox holds and no trade occurs. Therefore in practice search engines may have incentives to ensure their algorithms are not too precise.

3.5.2 Auctioning Prominence

This paper studies optimal search and pricing but avoids the question of how to sell off prominent advertising slots; several other papers focus on the latter but ignore the former. It would not be practical to formally introduce auctions into the current model, because there is no ex ante heterogeneity and everything is common knowledge.[18] Nevertheless

[17] The equivalent expression without prominence is $1/\left(1-x\right)h\left(0\right)$.

[18] The search engine could simply charge one fixed price for listing in g_1, and another fixed price for listing in g_2.

the model does suggest that per-click auctions may not perform well.

The theoretical auctions literature often assumes that a firm's value per-click is independent of its prominence (for example Athey and Ellison [5], Varian [73]). However this paper shows that this assumption is unlikely to hold. Indeed returning to the earlier Uniform and Normal examples, we see that the prominent firm has a *lower* value per click (see new Table). Intuitively less-prominent retailers charge high prices precisely because they receive fewer clicks, and a greater proportion of those clicks are from serious buyers. So although g_2 retailers earn less profits, they also get proportionally far fewer clicks. Consequently firms may be dissuaded from bidding aggressively in per-click auctions - 'winning' such an auction and becoming prominent can be very bad for profits. However more recent click-weighted auctions may perform better, because acquiring a prominent slot does not require such an aggressive bid.

	$d \sim U[-D, D]$	$d \sim N[0, 1]$
Prominent Firm		
Profit	$\frac{32}{75}D$	0.54
Clicks	1	1
Profit per click	$\frac{32}{75}D$	0.54
Non-prominent Firms		
Profit	$\frac{121}{375}D$	0.43
Clicks	$\frac{11}{25}$	0.45
Profit per click	$\frac{11}{15}D$	0.95

Finally, in principle one would like to understand how a profit-maximising platform

should organise links. Unfortunately this turns out be a complicated task.[19] Nevertheless one general principle (given our assumption of high valuations) is that a platform should force consumers to make extra clicks in order to learn prices. These small search frictions can substantially increase industry and prominent-group profits. Interestingly industry representatives can take advantage of the internet, set up a platform, and then use distortions in search order to improve industry-wide profits. Often this would involve setting $\tilde{s} = 0$ and therefore *not* making it hard for consumers to locate some of the firms.

3.5.3 Is Prominence Always Bad?

The general theme of this chapter is that prominence is bad for consumers and total welfare. I end with a simple counterexample, where consumer valuations are low. Precisely, consumers are uniformly distributed on a unit-length Hotelling line, with firms A and B at either end. A consumer located at $x \in [0,1]$ gets utility $V - tx - p_A$ when buying from firm A at price p_A, and $V - t(1-x) - p_B$ when buying from firm B at price p_B. Suppose that $V \in \left(\frac{3t}{2}, \frac{5t}{2}\right)$. Consumers know their x but cannot distinguish the firms without paying a search cost s and visiting one of their websites. When $s \equiv 0$ both firms charge a price t and the market is covered. In what follows, assume s is positive but $s \to 0$.

Without prominence: there is no equilibrium with full market coverage. If there

[19] Primarily this is because prices are only implicitly defined, though having only two groups is also quite restrictive in this context. Moreover in practice, some platforms may not enjoy much flexibility. For example the number of truly prominent positions on a webpage is limited, and if Google increased $\tilde{s}$, consumers would probably go elsewhere.

were, firms would charge $2t$, but then consumers with $x \approx \frac{1}{2}$ would anticipate this and not search. Nor is there a (symmetric) equilibrium with partial market coverage. If there were, given their expectations consumers located with $x \in [0, \tilde{x}] \cup [1 - \tilde{x}, 1]$ would search, for some $\tilde{x} \in \left(0, \frac{1}{2}\right)$. But then each retailer could charge slightly more than expected, and not lose any sales. Just as with a single-product Diamond Paradox, the market would break down.

With prominence: suppose the platform randomly displays a link to only one of the firms (and discards the other link). It is simple to show that there is an equilibrium in which every consumer searches, and the 'prominent' firm sets a price $\frac{V}{2t}$ or $V - t$ depending upon whether $V \lessgtr 2t$.

We have already shown that small search frictions lead to higher prices, and that prominence then causes further distortions. When valuations are low and search is random, each firm has too many Diamond consumers and too little thickness of demand. Just as in Chapter 2, this causes the market to unravel. When one firm is randomly made prominent, consumers do not know who is prominent, and therefore they reveal no match information for the firm to exploit. Price is lower and the market exists. Although this suggests we should be more cautious about our earlier results, it is also true that this example is very stylised, and it relies on quite specific values of V.

3.6 Conclusion

Although the internet has significantly reduced the costs of learning about firms, prominence and search order still seem to matter. Companies pay billions of dollars every year for online advertising, and small changes in the way search results are presented can have large impacts on demand. In this paper I build a simple model which explains why this is so. Most consumers do relatively little search, because they understand when they have found their best product. This means that less prominent firms have significantly less elastic demand curves, so they charge much higher prices and earn significantly lower profits. This happens even though the cost of making an extra click is negligible.

The paper has three interesting implications. Firstly, it explains why products in supermarket end-of-aisle displays tend to be much cheaper. Consumers catch sight of items whilst shopping. Products in obscure locations are largely seen by consumers who already intended to buy them (and hence have inelastic demand curves) whilst products in prominent displays are seen by almost everyone. Secondly, the growing empirical literature reconciles the large degree of online price dispersion by estimating search costs to be quite large. The theoretical model presented here shows that actually even small search costs can generate large distortions - and therefore suggests why current estimates of search costs may be exaggerated. Lastly, it is undoubtedly true that the internet has reduced search costs, but it has also made it incredibly easy for interested parties such as search engines to manipulate the order in which consumers search. Ironically therefore, instead of reducing price distortions, the internet could end up increasing them.

3.7 Appendix

3.7.1 Proofs of Search Rules

Proof of Lemma 10: Assume that conditional on visiting the $l+1^{th}$ website, the consumer visits every website until finding j. Then visiting the $l+1^{th}$ website gives payoff $v_j - p^* - \sum_{z=1}^{n-l} \frac{z\ s}{n-l} = v_j - p^* - s\frac{n-l+1}{2}$; this is better than buying i immediately if $v_i - p < v_j - p^* - s\frac{n-l+1}{2}$. The righthand side increases in l, so it is optimal to visit every website until finding j.

Suppose $v_i - p \geq v_j - p^* - s\frac{n-l+1}{2}$, and that after visiting the $l+1^{th}$ website, the consumer only clicks up to $k \leq n-l-1$ additional websites, and then if j has not been found, just buys i. Then visiting the $l+1^{th}$ website gives payoff $\frac{n-l-k-1}{n-l}\left(v_i - p - s\left(k+1\right)\right) + \frac{k+1}{n-l}\left(v_j - p^* - \sum_{z=1}^{k+1} \frac{z\ s}{k+1}\right)$; buying i immediately is better provided that $v_i - p \geq v_j - p^* - s\left[n-l-\frac{k}{2}\right]$, which holds because $v_i - p \geq v_j - p^* - s\frac{n-l+1}{2}$ and $k \leq n-l-1$. ■

The following proofs are written for any $\tilde{s} \in [0, D)$, so where necessary simply set $\tilde{s} = 0$.

Proof of Lemma 12: Part C follows from Lemma **11**. Consider Part B2ii), and assume that conditional on visiting the n_1+1^{th} website, the consumer searches until finding j. Searching firm n_1+1 gives expected payoff $v_j - p_2^E - \tilde{s} - s\frac{n_2+1}{2}$; this is better than buying i immediately if $v_i - p < v_j - p_2^E - \tilde{s} - s\frac{n_2+1}{2}$. From Part C, once n_1+1 is visited, the consumer does optimally search until finding j. Now suppose instead that $v_i - p \geq v_j - p_2^E - \tilde{s} - s\frac{n_2+1}{2}$, and that after visiting n_1+1, the consumer searches up to

$k \leq n_2 - 1$ additional websites and then, if j has not been found, buys i. Using a similar proof to that used in Lemma 10, buying i immediately (no further search) dominates visiting $n_1 + 1$.

Now consider B2, and assume $p_1^E < p_2^E + \tilde{s} - s\frac{n-l}{2}$. **Firstly**, if $v_i - p \geq v_j - p_2^E - \tilde{s} - s\frac{n_2+1}{2}$, the consumer prefers consuming i to clicking any links in g_2. Assume that after clicking on $l+1 \leq n_1$, he will click on links until j is found or every link in g_1 has been searched. Clicking on $l+1$ gives expected payoff $\frac{n_2}{n-l}\left(v_i - p - s\left(n_1 - l\right)\right) + \frac{n_1-l}{n-l}\left(v_j - p_1^E - \sum_{z=1}^{n_1-l}\frac{z\ s}{n_1-l}\right)$; this is better than consuming i immediately provided $v_i - p < v_j - p_1^E - s\left(\frac{n_1-l+1}{2} + n_2\right)$ [the righthand side increases in l, so clicking through to the end of g_1 is optimal]. Provided $p_1^E < p_2^E + \tilde{s} - s\frac{n-l}{2}$, this is also compatible with $v_i - p \geq v_j - p_2^E - \tilde{s} - s\frac{n_2+1}{2}$. **Secondly**, if $v_i - p < v_j - p_2^E - \tilde{s} - s\frac{n_2+1}{2}$, conditional on having searched g_1 and not found j, the consumer searches g_2 until finding j. Assume that conditional on clicking $l+1 < n_1$, the consumer searches through the entire list until finding j. Clicking $l+1$ gives payoff $v_j - \frac{n_1-l}{n-l}p_1^E - \frac{n_2}{n-l}\left(p_2^E + \tilde{s}\right) - \sum_{z=1}^{n-l}\frac{z\ s}{n-l}$; this is better than buying i immediately provided that $v_i - p < v_j - \frac{n_1-l}{n-l}p_1^E - \frac{n_2}{n-l}\left(p_2^E + \tilde{s}\right) - s\frac{n-l+1}{2}$. This holds since $v_i - p < v_j - p_2^E - \tilde{s} - s\frac{n_2+1}{2}$ and $p_1^E < p_2^E + \tilde{s} - s\frac{n-l}{2}$. Recall that $v_i - p < v_j - p_2^E - \tilde{s} - s\frac{n_2+1}{2}$ implies $v_i - p < v_j - p_1^E - s\left(\frac{n_1-l+1}{2} + n_2\right)$, so whenever $v_i - p < v_j - p_1^E - s\left(\frac{n_1-l+1}{2} + n_2\right)$, the consumer searches at least the whole of g_1 to find j.

Now consider B1. If $v_i - p \geq v_j - p_1^E - s\left(\frac{n_1-l+1}{2} + n_2\right)$, then $v_i - p \geq v_j - p_2^E - \tilde{s} - s\frac{n_2+1}{2}$ and the consumer never searches in g_2. Assume that after clicking on $l+1$, he visits up to

additional $k \leq n_1 - l - 1$ websites, and if j has not been found, he buys i. Searching $l+1$ gives expected payoff $\frac{n-l-k-1}{n-l}\left(v_i - p - s\left(k+1\right)\right) + \frac{k+1}{n-l}\left(v_j - p_1^E - \sum_{z=1}^{k+1} \frac{z\ s}{k+1}\right)$. Buying i immediately is better because $k \leq n_1 - l - 1$. ■

3.7.2 Proof of Proposition 13

From Step A (below) $p_2^E = \frac{n_2 - 1 + 2n_1 H\left(p_1^E - p_2^E - \tilde{s}\right)}{h(0)(n_2 - 1)}$ and Steps B and C, (i) $p_1^E = \frac{n_1 - 1 + 2n_2}{(n_1 - 1)h(0)}$ if $p_1^E < p_2^E + \tilde{s} + s\frac{n_2+1}{2} - D$ or (ii) $p_1^E = \frac{n_1 - 1 + 2n_2\left[1 - H\left(p_1^E - p_2^E - \tilde{s}\right)\right]}{(n_1 - 1)h(0) + 2n_2 h\left(p_1^E - p_2^E - \tilde{s}\right)}$ if $p_1^E \geq p_2^E + \tilde{s} + s\frac{n_2+1}{2} - D$. Notice that when $p_1^E < p_2^E + \tilde{s} + s\frac{n_2+1}{2} - D$ and $s \to 0$, then $p_2^E \to \frac{1}{h(0)} < p_1^E$. However by assumption $\tilde{s} < D$ and $s \to 0$, so $p_1^E < p_2^E + \tilde{s} + s\frac{n_2+1}{2} - D$ requires that $p_2^E > p_1^E$, yielding a contradiction.

Consequently any equilibrium has

$$p_2^E = \frac{n_2 - 1 + 2n_1 H\left(p_1^E - p_2^E - \tilde{s}\right)}{h\left(0\right)\left(n_2 - 1\right)} \text{ and } p_1^E = \frac{n_1 - 1 + 2n_2\left[1 - H\left(p_1^E - p_2^E - \tilde{s}\right)\right]}{\left(n_1 - 1\right)h\left(0\right) + 2n_2 h\left(p_1^E - p_2^E - \tilde{s}\right)}$$

I now demonstrate that there exists a unique solution to these two equations. Letting $\tilde{y} = p_1^E - p_2^E - \tilde{s}$,

$$\tilde{y} = \frac{n_1 - 1 + 2n_2\left[1 - H\left(\tilde{y}\right)\right]}{\left(n_1 - 1\right)h\left(0\right) + 2n_2 h\left(\tilde{y}\right)} - \frac{n_2 - 1 + 2n_1 H\left(\tilde{y}\right)}{h\left(0\right)\left(n_2 - 1\right)} - \tilde{s}$$

The lefthand side strictly increases in $\tilde{y}$. By assumption $\tilde{y} < 0$ so $h\left(\tilde{y}\right)$ increases in $\tilde{y}$, implying that the righthand side decreases in $\tilde{y}$. When $\tilde{y} = 0$, the lefthand side

exceeds the righthand side[20]; when $\tilde{y} = -D$ the lefthand side is less than the righthand side[21]. Therefore by continuity, there exists a unique $\tilde{y} \in (-D, 0)$ that solves the above equation. Given this $\tilde{y}$ there is then a unique equilibrium price vector $\left(p_1^E, p_2^E\right)$. The rest of the proof proceeds by considering Steps A-C.

STEP A: consider a firm in g_2. The firm has the k^{th} website with probability $\frac{1}{n_2}$ for any $k = n_1 + 1, ..., n$, and a mass $\frac{2}{n}$ of consumers like the firm's product. Suppose the firm is listed k^{th}. 1). with probability $\frac{n-k}{n-1}$, the consumer buys if $d \geq p_2 - p_2^E - s\frac{n+1-k}{2}$ since his other favourite is listed at some later website $h > k$. So with probability $\frac{n-k}{n-1}$ demand equals $1 - H\left(p_2 - p_2^E - s\frac{n+1-k}{2}\right)$. 2). with probability $\frac{n_1}{n-1}$ the consumer's other favourite is listed at some $h = 1, ..., n_1$; he only searches g_2 if $d < p_1^E - p_2^E - \tilde{s} - \frac{n_2+1}{2}s$, and once he's found what he's searching for, only buys it if $d < p_1^E - p_2$. So with probability $\frac{n_1}{n-1}$ demand equals $\min\left[H\left(p_1^E - p_2^E - \tilde{s} - \frac{n_2+1}{2}s\right), H\left(p_1^E - p_2\right)\right]$. 3). with probability $\frac{1}{n-1}$ the consumer's other favourite is listed at some $h = n_1 + 1, ..., k-1$; he only searches further if $d < -s\frac{n-h+1}{2}$, and once he's found what he's searching for, only buys it if $d < p_2^E - p_2$. This gives demand which we can write as $\frac{1}{n-1}\sum_{h=n_1+1}^{k-1} \min\left[H\left(-s\frac{n-h+1}{2}\right), H\left(p_2^E - p_2\right)\right]$.

The firm can be listed in any $k = n_1 + 1, ..., n$ so we must sum over k. 1). becomes $\sum_{k=n_1+1}^{n} \frac{n-k}{n-1}\left[1 - H\left(p_2 - p_2^E - s\frac{n+1-k}{2}\right)\right]$,

2). becomes $\frac{n_1 n_2}{n-1} \min\left[H\left(p_1^E - p_2^E - \tilde{s} - \frac{n_2+1}{2}s\right), H\left(p_1^E - p_2\right)\right]$.

3) is more complex and equals $\frac{1}{n-1}\sum_{k=n_1+1}^{n}\sum_{h=n_1+1}^{k-1} \min\left[H\left(-s\frac{n-h+1}{2}\right), H\left(p_2^E - p_2\right)\right]$. The latter simplifies to $\sum_{k=n_1+1}^{n} \frac{n-k}{n-1} \min\left[1 - H\left(s\frac{n-k+1}{2}\right), 1 - H\left(p_2 - p_2^E\right)\right]$. So ignor-

[20] The righthand side equals $\frac{n-1}{h(0)[n_1-1+2n_2]} - \frac{n-1}{h(0)(n_2-1)} - \tilde{s} < -\tilde{s} \leq 0$.

[21] The righthand side equals $\frac{n_1-1+2n_2}{(n_1-1)h(0)+2n_2h(-D)} - \frac{1}{h(0)} - \tilde{s} \geq \frac{n_1-1+2n_2}{h(0)[n_1-1+2n_2]} - \frac{1}{h(0)} - \tilde{s} = -\tilde{s} > -D$.

ing constants, demand equals

$$\begin{aligned} D_2 &= \sum_{k=n_1+1}^{n} (n-k)\left[1-H\left(p_2-p_2^E-s\frac{n+1-k}{2}\right)\right] \\ &\quad +n_1n_2\min\left[H\left(p_1^E-p_2^E-\tilde{s}-\frac{n_2+1}{2}s\right),H\left(p_1^E-p_2\right)\right] \\ &\quad +\sum_{k=n_1+1}^{n}(n-k)\min\left[1-H\left(s\frac{n-k+1}{2}\right),1-H\left(p_2-p_2^E\right)\right] \end{aligned}$$

As $s\to 0$, equilibrium demand is $\sum_{k=n_1+1}^{n}(n-k)+n_1n_2H\left(p_1^E-p_2^E-\tilde{s}\right)$ or $\frac{n_2(n_2-1)}{2}+n_1n_2H\left(p_1^E-p_2^E-\tilde{s}\right)$. Equilibrium thickness of demand becomes $h(0)\sum_{k=n_1+1}^{n}(n-k)$ or $h(0)\frac{n_2(n_2-1)}{2}$. So

$$p_2^E=\frac{n_2-1+2n_1H\left(p_1^E-p_2^E-\tilde{s}\right)}{h(0)(n_2-1)}$$

For $p_2\le p_2^E+\tilde{s}+s\frac{n_2+1}{2}$, profit is concave in p_2 and therefore maximised at $p_2=p_2^E$ provided $\tilde{s}$ is small enough. This is because profit is a sum of segments, each of which is concave in p_2.[22] For any $p_2\ge p_2^E+\tilde{s}+s\frac{n_2+1}{2}$, profit is decreasing in p_2. This is because the first derivative of profit with respect to p_2 is a sum of negative terms.[23]

[22] 1). Consider terms of the form $p_2\left[1-H\left(p_2-p_2^E-\tilde{s}-s\frac{n+1-k}{2}\right)\right]$. When p_2 is small, demand equals 1. For larger p_2, the second derivative is strictly negative - differentiating twice gives $-2h\left(p_2-p_2^E-\tilde{s}-s\frac{n+1-k}{2}\right)-p_2h'\left(p_2-p_2^E-\tilde{s}-s\frac{n+1-k}{2}\right)$, where $h'\left(p_2-p_2^E-\tilde{s}-s\frac{n+1-k}{2}\right)\gtrless 0$ since $p_2-p_2^E-\tilde{s}-s\frac{n+1-k}{2}\le s\frac{k-n_1}{2}\approx 0$, and $h(d)$ is symmetric and logconcave.

2). Terms of the form $p_2\min\left[H\left(p_1^E-p_2^E-\tilde{s}-\frac{n_2+1}{2}s\right),H\left(p_1^E-p_2\right)\right]$ are linear in p_2 over the considered interval.

3). Terms of the form $p_2\min\left[1-H\left(s\frac{n-k+1}{2}\right),1-H\left(p_2-p_2^E\right)\right]$ are linear in p_2 for low p_2. For larger p_2, the second derivative is $-2h\left(p_2-p_2^E\right)-p_2h'\left(p_2-p_2^E\right)$, which is negative if $\tilde{s}=0$ since $s\to 0$ and $h(d)$ is symmetric logconcave, and is also negative provided $\tilde{s}$ is not too large.

[23] We need to show that terms of the form 1). $1-H\left(p_2-p_2^E-\tilde{s}-s\frac{n+1-k}{2}\right)-p_2h\left(p_2-p_2^E-\tilde{s}-s\frac{n+1-k}{2}\right)$, 2). $1-H\left(p_2-p_1^E\right)-p_2h\left(p_2-p_1^E\right)$ and 3). $1-H\left(p_2-p_2^E\right)-p_2h\left(p_2-p_2^E\right)$ are negative.

Consider $p_2\left[1-H\left(p_2-z\right)\right]$. This is logconcave in p_2 and therefore has a unique maximum given by $p_2^*=\max\left[z-D,\frac{1-H(p_2^*-z)}{h\left(p_2^*-z\right)}\right]$. It is simple to show that provided $z>\frac{1}{2h(0)}$, $p_2^*<z$ and therefore that

STEP B: consider a firm in g_1, and look for an interior solution. The firm has the k^{th} website with probability $\frac{1}{n_1}$ for any $k = 1, ..., n_1$, and a mass $\frac{2}{n}$ of consumers like the firm's product. Suppose the firm is listed k^{th}. 1). with probability $\frac{n_1-k}{n-1}$ the consumer's other favourite is listed at some $h = k+1, ..., n_1$, so he buys if $d \geq p - p_1^E - s\left(\frac{n_1-k+1}{2} + n_2\right)$. So with probability $\frac{n_1-k}{n-1}$, demand equals $1 - H\left(p - p_1^E - s\left(\frac{n_1-k+1}{2} + n_2\right)\right)$. 2). with probability $\frac{n_2}{n-1}$ the consumer's other favourite is listed at some $h = n_1 + 1, ..., n$, so he buys if $d \geq p - p_2^E - \tilde{s} - s\frac{n_2+1}{2}$. So with probability $\frac{n_2}{n-1}$, demand equals $1 - H\left(p - p_2^E - \tilde{s} - s\frac{n_2+1}{2}\right)$. 3). with probability $\frac{1}{n-1}$ the consumer's other favourite is listed at any $h = 1, ..., k-1$; he only searches further if $d < -s\left(\frac{n_1-h+1}{2} + n_2\right)$, and once he's found what he's searching for, only buys it if $d < p_1^E - p$. This contributes demand of $\frac{1}{n-1}\sum_{h=1}^{k-1} \min\left[H\left(-s\left(\frac{n_1-h+1}{2} + n_2\right)\right), H\left(p_1^E - p\right)\right]$. As before, we need to sum over k, which (ignoring constants) gives the following expression for demand:

$$\begin{aligned} D_1 &= \sum_{k=1}^{n_1} (n_1 - k)\left[1 - H\left(p - p_1^E - s\left(\frac{n_1 - k + 1}{2} + n_2\right)\right)\right] \\ &\quad + n_1 n_2\left[1 - H\left(p - p_2^E - \tilde{s} - s\frac{n_2 + 1}{2}\right)\right] \\ &\quad + \sum_{k=1}^{n_1} (n_1 - k) \min\left[H\left(-s\left(\frac{n_1 - k + 1}{2} + n_2\right)\right), H\left(p_1^E - p\right)\right] \end{aligned}$$

As $s \to 0$, equilibrium demand becomes $\sum_{k=1}^{n_1}(n_1 - k) + n_1 n_2\left[1 - H\left(p_1^E - p_2^E - \tilde{s}\right)\right]$ or $n_1\left[\frac{n_1-1}{2} + n_2\left[1 - H\left(p_1^E - p_2^E - \tilde{s}\right)\right]\right]$.Equilibrium thickness of demand is $h(0)\sum_{k=1}^{n_1}(n_1 - k)$

for any $p_2 \geq z$, $1 - H(p_2 - z) - p_2 h(p_2 - z) < 0$. Since in each of the segments we have $z > \frac{1}{2h(0)}$, we are done.

$+n_1 n_2 h\left(p_1^E - p_2^E - \tilde{s}\right)$ or $n_1\left[\frac{n_1-1}{2}h\left(0\right) + n_2 h\left(p_1^E - p_2^E - \tilde{s}\right)\right]$. Hence equilibrium price is

$$p_1^E = \frac{\frac{n_1-1}{2} + n_2\left[1 - H\left(p_1^E - p_2^E - \tilde{s}\right)\right]}{\frac{n_1-1}{2}h\left(0\right) + n_2 h\left(p_1^E - p_2^E - \tilde{s}\right)}$$

We again need to check that the first order condition is sufficient to characterise the equilibrium. Using a similar method to before, we can show the $p_1 D_1$ is concave in p_1 for $p_1 \leq p_1^E + s\frac{n_2+1}{2}$. For $p_1 > p_1^E + s\frac{n_2+1}{2}$ it is again sufficient to show that each of the profit function segments has a negative first derivative. The segment $p_1\left[1 - H\left(p_1 - p_2^E - \tilde{s} - s\frac{n_2+1}{2}\right)\right]$ is the only part that requires more attention. $p_1\left[1 - H\left(p_1 - p_2^E - \tilde{s} - s\frac{n_2+1}{2}\right)\right]$ is logconcave and $p_1^* = \max\left[p_2^E + \tilde{s} - D, \frac{1-H\left(p_1^* - p_2^E - \tilde{s} - s\frac{n_2+1}{2}\right)}{h\left(p_1^* - p_2^E - \tilde{s} - s\frac{n_2+1}{2}\right)}\right]$ maximises it. Provided that the solution to $p_1^* = \frac{1-H\left(p_1^* - p_2^E - \tilde{s} - s\frac{n_2+1}{2}\right)}{h\left(p_1^* - p_2^E - \tilde{s} - s\frac{n_2+1}{2}\right)}$ is below p_1^E then for $p_1 > p_1^E + s\frac{n_2+1}{2}$, the derivative of $p_1\left[1 - H\left(p_1 - p_2^E - \tilde{s} - s\frac{n_2+1}{2}\right)\right]$ is negative and the first order condition is sufficient.

STEP C: consider a firm in g_1, and look for a corner solution in which $p_1^E < p_2^E + \tilde{s} + s\frac{n_2+1}{2} - D$. Modifying the demand function given in Step B, one can show that this implies an equilibrium price $p_1^E = \frac{\frac{n_1-1}{2} + n_2}{\frac{n_1-1}{2}h(0)}$. ■

3.7.3 Other Proofs

Proof of Proposition 16: Suppose $p_1^E < \frac{1}{h(0)}$, and let $\tilde{y} = p_1^E - p_2^E - \tilde{s}$. A g_1 firm gets demand $\frac{2}{n}\left[\frac{n_1-1}{n-1}\frac{1}{2} + \frac{n_2}{n-1}\left[1 - H\left(\tilde{y}\right)\right]\right]$, and a g_2 firm gets demand $\frac{2}{n}\left[\frac{n_2-1}{n-1}\frac{1}{2} + \frac{n_1}{n-1}H\left(\tilde{y}\right)\right]$. So industry profit equals

$$
\begin{aligned}
& \tfrac{1}{n(n-1)}\left[\, n_1 p_1^E\left[n_1-1+2n_2\left[1-H\left(\tilde{y}\right)\right]\right]+n_2 p_2^E\left[n_2-1+2n_1 H\left(\tilde{y}\right)\right]\,\right] \\
> \;& \tfrac{1}{n(n-1)}\left[\, n_1 p_1^E\left[n_1-1+2n_2\right]+n_2 p_2^E\left[n_2-1\right]\,\right] \\
= \;& \tfrac{1}{n(n-1)}\left[\, n_1\left[\tfrac{n_1-1+2n_2[1-H(\tilde{y})]}{(n_1-1)h(0)+2n_2h(\tilde{y})}\right]\left[n_1-1+2n_2\right]+n_2\left[\tfrac{n_2-1+2n_1H(\tilde{y})}{(n_2-1)h(0)}\right]\left[n_2-1\right]\,\right] \\
\geq \;& \tfrac{1}{n(n-1)}\left[\, n_1\left[\tfrac{n_1-1+2n_2[1-H(\tilde{y})]}{h(0)[n_1-1+2n_2]}\right]\left[n_1-1+2n_2\right]+n_2\left[\tfrac{n_2-1+2n_1H(\tilde{y})}{(n_2-1)h(0)}\right]\left[n_2-1\right]\,\right] \\
= \;& \tfrac{1}{n(n-1)}\tfrac{1}{h(0)}\left[n_1\left(n_1-1\right)+n_2\left(n_2-1\right)+2n_1n_2\right] \\
= \;& \tfrac{1}{h(0)} = \text{Profit in the 'without prominence' case. } \blacksquare
\end{aligned}
$$

Proof of Proposition 18:

$$
\begin{aligned}
\text{Profit} \;=\;& \tfrac{1}{n(n-1)}\left[\, n_1\tfrac{\left[n_1-1+2n_2[1-H(\tilde{y})]\right]^2}{(n_1-1)h(0)+2n_2h(\tilde{y})}+n_2\tfrac{\left[n_2-1+2n_1H(\tilde{y})\right]^2}{(n_2-1)h(0)}\,\right] \\
\propto\;& n_1\tfrac{\left[n_1-1+2n_2[1-H(\tilde{y})]\right]^2}{(n_1-1)h(0)+2n_2h(\tilde{y})}+n_2\tfrac{\left[n_2-1+2n_1H(\tilde{y})\right]^2}{(n_2-1)h(0)} \\
\tfrac{\partial\ \text{Profit}}{\partial\tilde{y}} \propto\;& -4n_1n_2h\left(\tilde{y}\right)p_1^E+4n_1n_2h\left(\tilde{y}\right)p_2^E-2n_1n_2h'\left(\tilde{y}\right)\left[p_1^E\right]^2 \\
\propto\;& -\left[\, p_1^E-p_2^E+\tfrac{1}{2}\left[p_1^E\right]^2\tfrac{h'(\tilde{y})}{h(\tilde{y})}\right] \\
\tfrac{\partial\ \text{Profit}}{\partial\tilde{s}} \propto\;& p_1^E-p_2^E+\tfrac{1}{2}p_1^Ep_1^E\tfrac{h'(\tilde{y})}{h(\tilde{y})}\ \text{since}\ \tfrac{\partial\tilde{y}}{\partial\tilde{s}}<0. \ \blacksquare
\end{aligned}
$$

Proof of Lemma 19: Total surplus equals

$$
\begin{aligned}
TS \;=\;& \frac{n_1\left(n_1-1\right)}{n\left(n-1\right)}\max\left(v_i,v_j\right)+\frac{n_2\left(n_2-1\right)}{n\left(n-1\right)}\left[\max\left(v_i,v_j\right)-\tilde{s}\right] \\
& +\frac{2n_1n_2}{n\left(n-1\right)}\left[\,\int_{\tilde{y}}^{D}h\left(z\right)v_1\left(z\right)dz+\int_{-D}^{\tilde{y}}h\left(z\right)\left[\,v_2\left(z\right)-\tilde{s}\right]dz\right] \\
\frac{\partial TS}{\partial\tilde{s}} \;=\;& -\frac{n_2\left(n_2-1\right)}{n\left(n-1\right)}-\frac{2n_1n_2}{n\left(n-1\right)}H\left(\tilde{y}\right)+\frac{2n_1n_2}{n\left(n-1\right)}\frac{\partial\tilde{y}}{\partial\tilde{s}}h\left(\tilde{y}\right)\left[v_2\left(\tilde{y}\right)-v_1\left(\tilde{y}\right)-\tilde{s}\right]
\end{aligned}
$$

Note that $\frac{\partial\tilde{y}}{\partial\tilde{s}}<0$ so a sufficient condition for $\frac{\partial TS}{\partial\tilde{s}}<0$ is that $v_2\left(\tilde{y}\right)-v_1\left(\tilde{y}\right)-\tilde{s}\geq 0$.

This requires that $p_2^E-p_1^E\geq 0$ since $v_1\left(\tilde{y}\right)-p_1^E=v_2\left(\tilde{y}\right)-p_2^E-\tilde{s}$. ■

Chapter 4

Small Switching Costs are Pro-Competitive

Abstract: When consumers face switching costs, they are less likely to change supplier in order to get a better deal. On the one hand firms want to exploit this, and charge their existing customers a high price. But on the other hand firms want to charge a low price and attract people who can then be exploited in the future. The general consensus in the theoretical literature is that switching costs are anti-competitive because the former effect dominates the latter. However these models are often based on the assumptions that products are homogeneous and/or switching costs are very large. I build a less restrictive model, and find that switching costs are strongly pro-competitive and often beneficial to consumers.

4.1 Introduction

There are a number of reasons why consumers might bear a cost when switching suppliers, and these costs are often estimated to be significant. Some switching costs are physical: in order to change savings accounts, for example, a consumer must spend time and effort visiting a bank and filling out forms. Other switching costs are psychological: consumers may gradually become attached to a particular product or brand, and (for no good economic reason) be reluctant to try anything else. Switching costs may also be articifically-created: airmiles and supermarket loyalty points are two obvious examples. Klemperer [45] discusses many other interesting ways in which switching costs might arise. But irrespective of where they come from, these switching costs can be surprisingly large. For example Dubé *et al* [22] estimate that switching costs for margarine and orange juice are between 15% and 19% of the retail price. Meanwhile using (pre-2000) survey data for the UK gas market, Giulietti *et al* [33] find that 55% of consumers would stay loyal to British Gas even if switching elsewhere saved £8 per month.

Consumers' reluctance to change supplier can give rise to an interesting tension. Since switching costs discourage consumers from changing supplier, this gives a firm some ex post market power over its existing customers. It would therefore be tempting to conclude that switching costs enable firms to charge a high price and to therefore earn high profits - much like search costs did in the previous Chapter. However by their very nature, switching costs imply some notion of dynamic competition. In particular, if a firm really can exploit its customers and earn high profits off them - then market share

is a valuable asset, and firms should compete aggressively for it. In particular ex ante competition for the right to exploit consumers ex post, could be very fierce. The impact of switching costs on market competition is therefore ambiguous.

Nevertheless the dominant view within the theoretical literature is that switching costs are anti-competitive. Klemperer [45] for example writes (Page 516): "We find a presumption that firms' incentives to exploit repeat purchasers dominate their incentives to attract new customers, and so lead to higher prices in markets with switching costs." However many of the more influential models assume either undifferentiated products (Farrell and Shapiro [26] and Padilla [58]) or switching costs that are so large that nobody ever switches (Beggs and Klemperer [10]). In reality many consumers do switch suppliers even in mature markets with stable relative prices, and very few products are entirely homogeneous.

In this paper I present a framework in which two retailers sell differentiated products, and switching costs are small but not insignificant. There are overlapping generations of consumers who each live for two periods, and who have tastes which are independent across time. I also look at three different scenarios which vary how much pricing freedom the two retailers have. The major result is that small switching costs are always pro-competitive and bad for profits, and under a very reasonable condition are good for consumers. The paper makes three other contributions. Firstly, it shows that in an 'infinite-horizon model', even an incumbent monopolist facing a brand new entrant (which has no installed base) may charge less than in the absence of switching costs. Secondly, it

shows that sometimes in a 'two-period model' prices may follow a 'ripoffs-then-bargains' pattern rather than the more standard 'bargains-then-ripoffs'. And thirdly it shows that when consumers have switching costs, firms are often made worse off by the ability to price discriminate based upon consumers' past purchase behaviour.

4.2 Literature

There is now a large body of work on switching costs, and Klemperer [45] and Farrell and Klemperer [25] provide excellent surveys. This literature review is therefore brief and focuses on three important classes of model.

The seminal papers Klemperer [42] (with homogeneous goods) and [43] (with differentiated products) focused on **two-period models** of switching costs. In the first period consumers are unattached, and firms invest in future market share by charging low prices. In the second period consumers are attached and face switching costs - so firms exploit them by charging high prices.[1] Prices therefore follow a pattern of 'bargains-then-ripoffs'. The net effect is ambiguous - but these early papers lean towards the view that switching costs are anti-competitive. The model in Section 4.4 takes a special case of the Klemperer [43] model (namely $\mu + \nu = 1$), but generalises it to arbitrary discount factors and places less restrictions on consumer preferences. I show that with the latter generalisation, prices may be lower *in both periods* and follow

[1] I will interpret these two-period models as consisting of overlapping generations of consumers (who live for two periods), and infinitely-lived firms who can price discriminate between old and young consumers.

a 'ripoffs-then-bargains' pattern. Small switching costs are *always* pro-competitive, and since I allow firms and consumers to have different discount factors, the intuition behind this result is particularly easy to understand.

Another strand of the literature looks at the interaction of switching costs and **behaviour-based price discrimination**. In Chen [16] consumers live for two periods and have heterogeneous switching costs, whilst firms sell homogeneous products. In the second period each firm charges two prices - a high price to exploit its own customers, and a low price to 'poach' some (low switching cost) customers from the rival. Firms make strictly positive profits and therefore benefit from the existence of positive switching costs. Interestingly *proportional increases* in switching costs are good for profits whilst *lump-sum increases* are bad (Chen [16], Gehrig and Stenbacka [31], Bouckaert *et al* [13]). In a model with differentiated products, Cabral [15] demonstrates that starting from zero, a small increase in the switching cost reduces average price. Section 4.5 presents a similar model, but demonstrates that profits are always monotonically decreasing in the switching cost. I also compare profits, prices and welfare with the two-period model (where firms cannot use behaviour-based discrimination) and find results that differ from Chen [16].[2]

Finally in **infinite-horizon models**, firms face overlapping generations of consumers but charge one price in each time period - and are therefore unable to price discriminate. They avoid the 'end of time' effects that exist in two-period models, but are substantially

[2]Fudenberg and Tirole [29] and Villas-Boas [78], [79] study behaviour-based price discrimination in the absence of switching costs.

more difficult to analyse. Early contributions include Farrell and Shapiro [26], Beggs and Klemperer [10], Padilla [58] and To [72]. All four papers argue that switching costs unambiguously raise prices. However Beggs and Klemperer [10] and To [72] both assume that switching costs are so large that no 'old' consumer ever changes supplier. The model by Farrell and Shapiro [26] has the unusual features that firms price sequentially and alternate between 'harvesting' and 'investing'. Each period one firm (the incumbent) acts as price-leader and sells to every old consumer, whilst the other firm (the entrant) acts as price-follower and sells to every young consumer; the roles of incumbent and entrant are then reversed in the next period. Importantly, no switching is ever observed in equilibrium. Finally firms in Padilla's [58] model sell homogeneous products and equilibrium is in mixed strategies; he again finds that in equilibrium no old consumer finds it optimal to switch supplier.

The earlier infinite-horizon models therefore all share the property that - either by assumption or in equilibrium - no consumer ever switches supplier. In reality of course many consumers do engage in switching, and this has motivated more recent work by Doganoglu [21], Dubé *et al* [22] and Viard [77]. Doganoglu [21] models experience goods, and shows that starting from zero, a small increase in the switching cost causes steady state price to fall. Dubé *et al* [22] (in a discrete-choice framework) and Viard [76] (with firms located on a Hotelling line) show numerically that switching costs have ambiguous effects on steady state prices. In Section 4.6 I build a model where prices are reduced by switching costs - and I prove analytically that the result holds for a wide range of

parameters. Insights from my other models also enable me to provide a clearer intuition.

4.3 Assumptions

There are two infinitely-lived firms called A and B, which sell differentiated substitute products. Consumers live for two periods, and a unit mass of them is born each period. Therefore at any point in time, there are equal numbers of 'young' and 'old' consumers.

A consumer living in period t values the two products at V_A^t and V_B^t respectively - where these valuations are large enough that in equilibrium both young and old consumers make a purchase. The difference in valuations $d^t = V_B^t - V_A^t$ is a random variable with support $[-D, D]$ and a distribution function $H(d)$. The density function $h(d)$ is symmetric, and the associated hazard rate is increasing. Furthermore, d^t and d^{t+1} are assumed to be independent - so when a consumer becomes old, they receive a brand new draw from the 'taste' distribution which is independent of their preference when young. Firms and consumers are risk-neutral, and have discount factors δ_f and δ_c respectively which both lie in $(0, 1)$. Finally, if an old consumer bought from firm i when they were young, but now wants to buy from firm $j \neq i$, a switching cost $s > 0$ must be paid.

A young consumer born in period t learns d^t but not d^{t+1}. He observes period-t prices, and uses these to forecast prices in the next period. The young consumer chooses which product to purchase, in order to maximise his expected discounted lifetime surplus. When period $t+1$ arrives and the consumer is now old, he observes prices and d^{t+1}. He then decides whether to stick with his old supplier, or pay the switching cost and buy

from the rival.

The two retailers choose prices in order to maximise lifetime discounted profits. Demand comes from three sources: unattached young consumers, old consumers attached to firm A, and old consumers attached to firm B. In Section 4.5 retailers can charge different prices for each of the three groups of consumer. In Section 4.4 firms can only discriminate between young and old consumers, whilst Section 4.6 studies the case where no discrimination is possible.

4.3.1 Three Effects on Price

I begin by briefly reviewing three ways in which switching costs affect prices:

1. **Harvesting**

2. **Investing**

3. **Elasticity effect**

Firstly switching costs discourage a consumer from changing supplier. Retailers therefore have incentives to **harvest** previous customers by charging them a *high* price. However a firm can only harvest consumers in the future if it locks them in during the present. Retailers therefore have incentives to **invest** in market share (and therefore future profits) by charging a *low* price. However if one firm invests heavily in market share, it makes its rival more aggressive in the future; this reduces incentives to invest in market share (Klemperer [45]). Lastly the **elasticity effect** arises because consumers

understand that a low-price retailer will probably increase its current market share, and then charge a high price in the next period. This makes it less attractive to buy from a low-price retailer - consumer demand become less elastic and ceteris paribus firms respond with a *high* price.

The literature usually argues that the investing effect is weak, and is dominated by the harvesting and consumer elasticity effects. I argue instead that when switching costs are small, the investing effect is dominant, and switching costs therefore reduce prices and profits.

4.4 Partial Discrimination

Firms can observe whether a consumer is young (and has no product attachment) or old - but cannot price discriminate based upon who an old consumer previously bought from. To simplify matters I look at the prices charged to consumers born in time 1, and drop time superscripts whenever no confusion is likely to arise. The two retailers charge $p_{A,y}$ and $p_{B,y}$ respectively to young consumers, and then $p_{A,o}$ and $p_{B,o}$ to old consumers. Young consumers buy from firm A if $d^1 \leq \tilde{y}$ otherwise they buy from firm B.

4.4.1 Solving for Prices

I solve the model by backwards induction and therefore begin by studying how firms set prices for old consumers.

Prices Paid by Old Consumers

The purchase decision of an old consumer is characterised as follows. An old consumer who previously bought from firm A will stay with A provided that $d^2 \leq p_{B,o} - p_{A,o} + s$ - or in words, provided that the (utility) benefits of switching are less than the costs. Similarly an old consumer who previously bought from firm B will only switch to firm A if $d^2 \leq p_{B,o} - p_{A,o} - s$. Therefore letting $D^2_{i,o}$ be retailer i's second-period demand:

$$D^2_{A,o} = 1 - D^2_{B,o} = H(\tilde{y})H(p^2_{B,o} - p^2_{A,o} + s) + [1 - H(\tilde{y})]\, H(p^2_{B,o} - p^2_{A,o} - s)$$

The two interior first order conditions can be written as:

$$p^2_{A,o} = \frac{H(\tilde{y})H(p^2_{B,o} - p^2_{A,o} + s) + [1 - H(\tilde{y})]\, H(p^2_{B,o} - p^2_{A,o} - s)}{H(\tilde{y})h(p^2_{B,o} - p^2_{A,o} + s) + [1 - H(\tilde{y})]\, h(p^2_{B,o} - p^2_{A,o} - s)} \tag{4.1}$$

$$p^2_{B,o} = \frac{1 - H(\tilde{y})H(p^2_{B,o} - p^2_{A,o} + s) - [1 - H(\tilde{y})]\, H(p^2_{B,o} - p^2_{A,o} - s)}{H(\tilde{y})h(p^2_{B,o} - p^2_{A,o} + s) + [1 - H(\tilde{y})]\, h(p^2_{B,o} - p^2_{A,o} - s)} \tag{4.2}$$

Proving existence and uniqueness of equilibrium is much harder than in a standard problem with no switching cost. Ordinarily logconcavity of $1 - H(d)$ would be sufficient. The problem here is that demand is a mixture of two different distributions - and whilst individually they are logconcave, the mixture need not be. However under some circumstances one can show that for any $\tilde{y}$, the first order conditions (4.1) and (4.2) are sufficient, and implicitly define a unique pair of equilibrium prices. Sufficient conditions

for this are 1) $h(d)$ is close enough to being uniform and $\frac{s}{D} < \frac{3}{5}$ or 2) s is small enough.[3]

In the equilibrium of the entire game, retailers will each sell to half of the young consumers, so $\tilde{y} = 0$. Substituting this into (4.1) and (4.2) and looking for a symmetric equilibrium, we find that old consumers pay a price

$$p_{A,o}^2 = p_{B,o}^2 = \frac{1}{2h(s)} \tag{4.3}$$

Switching costs create two margins of competition. When firm A lowers its price, it encourages some marginal consumers to switch from B to itself, and it also persuades some marginal consumers not to switch from itself to B. These groups of marginal consumers have $d^2 = -s$ and $d^2 = s$ respectively, and their collective mass is equal to $h(s)$. Therefore when $h(s)$ is larger, more marginal consumers change their behaviour following small price cuts - so competition is fiercer and prices lower.

Remark 20 *An increase in the switching cost will* reduce *prices paid by old consumers when $h(d)$ is U-shaped*

Surprisingly, switching costs can reduce second-period prices - a point not made elsewhere in the literature. Intuitively a higher switching cost implies that marginal consumers (those just indifferent about switching suppliers) have more extreme tastes. When the distribution of preferences is U-shaped, relatively more consumers have these

[3] If $|h'(d)|$ is too large then a firm would deviate from the prices that solve (4.1) and (4.2). For example if $h(d)$ is strongly U-shaped, deviating and charging a low price may be profitable since demand increases rapidly. Similarly when s is large, a firm may profitably deviate by increasing price and serving only its locked consumers.

extreme tastes. This means that small price cuts are more successful at winning new business, and this encourages retailers to price more aggressively. When d is uniformly distributed, price is constant in the switching cost; any small deviation from a uniform distribution results in prices either increasing or decreasing in s.

Finally, it is important to understand how first-period market share affects second-period outcomes. If $\Pi_{A,o}$ is A's profit on old consumers, then

Lemma 21 *In a symmetric equilibrium*

$$\begin{aligned}\frac{\partial p_{A,o}}{\partial \tilde{y}} &= -\frac{\partial p_{B,o}}{\partial \tilde{y}} = \frac{h(0)\left[2H(s)-1\right]}{3h(s)} \\ \frac{\partial \Pi_{A,o}}{\partial \tilde{y}} &= \underbrace{\frac{h(0)\left[2H(s)-1\right]}{6h(s)}}_{\textit{Higher price per buyer}} + \underbrace{\underbrace{\frac{h(0)\left[2H(s)-1\right]}{2h(s)}}_{\textit{Lock-in effect}} - \underbrace{\frac{h(0)\left[2H(s)-1\right]}{3h(s)}}_{\textit{More aggressive rival}}}_{\textit{Total impact on (value of) market share}} > 0\end{aligned}$$

A firm with a larger market share charges a higher price and earns more profit. Moreover since $H(s)$ is logconcave, these effects are stronger when s is larger. The relationship between market share and price is well-understood (see for example Klemperer [45]). Each retailer wants to exploit its locked consumers by charging a high price, but also cater for consumers who are locked to its rival by charging a low price. The benefits of charging a high price are greater for retailers with a larger market share. The relationship between first-period market share and second-period profits depends upon the following factors:

1. Firm A earns a higher price on each customer[4]

2. Firm A has a higher share of the mature market

   (a) **Lock-in effect:** Young consumers tend to buy from their initial supplier rather than switch. So a higher share of young consumers translates, ceteris paribus, into a higher share of old consumers.

   (b) **More aggressive rival:** A charges more while B charges less. This has a negative effect on A's demand

It has been suggested in the literature that the 'more aggressive rival' effect could be so strong that mature-market profits are *decreasing* in young-market share. This cannot happen in the current model. A small increase in $\tilde{y}$ leaves the identity of marginal consumers - and therefore the thickness of demand - unchanged. Hence a firm increases its price if and only if its demand increases. So if higher $\tilde{y}$ lowers A's demand, it must also reduce A's price and increase B's price. However A then faces a less aggressive rival and had more locked in consumers, so its demand could not possibly have fallen - a contradiction which proves that $p^2_{A,o}$, $D^2_{A,o}$ and $\Pi^2_{A,o}$ all increase.[5]

[4] Note that A has demand of $\frac{1}{2}$ and increases its price by $\frac{h(0)[2H(s)-1]}{3h(s)}$.

[5] Outside of symmetric equilibrium, this monotone relationship between $\tilde{y}$ and $p^2_{A,o}$ and $D^2_{A,o}$ holds provided $h(d)$ is sufficiently close to being uniform.

Prices Paid by Young Consumers

Young consumers buy from firm A if their match difference d^1 is below a threshold $\tilde{y}$, and buy from firm B otherwise. Since consumers know $\tilde{y}$, they can forecast the prices they will face when old (namely $p_{A,o}$ and $p_{B,o}$). For convenience the difference in these prices is written $T = p_{B,o} - p_{A,o}$. A young consumer who buys from firm A enjoys an expected lifetime payoff of:

$$V_{A,y}^1 - p_{A,y}^1 + \delta_c \left[E_d V_A^2 - p_{A,o} + \int_{T+s}^{D} h(z) [z - T - s] \, dz \right]$$

This is composed of a certain payoff when young of $V_{A,y}^1 - p_{A,y}^1$, plus an expected payoff when old of $E_d V_A^2 - p_{A,o} + \int_{T+s}^{D} h(z) [z - T - s] \, dz$ which is discounted. Intuitively an old consumer can always stick with firm A and get an (ex ante) expected payoff $E_d V_A^2 - p_{A,o}$. However when his realised preference for B's product is strong enough (namely when $d^2 \geq T + s$) he switches supplier and earns some additional surplus. Similarly the expected lifetime payoff of a young consumer who buys from firm B is:

$$V_{B,y}^1 - p_{B,y} + \delta_c \left[E_d V_B^2 - p_{B,o} + \int_{-D}^{T-s} h(z) [-z + T - s] \, dz \right]$$

Young consumers with $d^1 = \tilde{y}$ are indifferent about buying from A or B. This threshold $\tilde{y}$ is therefore implicitly defined by the equation[6]:

$$\tilde{y} = p_{B,y} - p_{A,y} + \delta_c \left[T + \int_{T+s}^{D} h(z) [z - T - s] \, dz - \int_{-D}^{T-s} h(z) [-z + T - s] \, dz \right]$$

It is important to know how small changes in a retailer's price affect $\tilde{y}$. It is well-understood within the literature that a positive switching cost makes demand less elastic amongst young consumers (see for example Klemperer [45]). A small reduction in $p_{A,y}$ makes firm A more attractive to consumers and therefore causes $\tilde{y}$ to increase. But $\tilde{y}$ increases less than one-for-one, since consumers anticipate that A will price higher in the mature market.

Lemma 22 *In a symmetric equilibrium*

$$\frac{\partial \tilde{y}}{\partial p_{A,y}} = \frac{-1}{1 + \delta_c \cdot \underbrace{[2H(s) - 1]}_{\text{Probability term}} \cdot \underbrace{2 \dfrac{h(0)[2H(s) - 1]}{3h(s)}}_{\text{Price term}}} > -1 \tag{4.4}$$

Equation (4.4) can be understood as follows. In a symmetric equilibrium, all old consumers with $d^2 < -s$ buy from firm A and all those with $d^2 > s$ buy from firm B. It is only when $d^2 \in [-s, s]$ that an old consumer's choice of retailer is affected by the choice they made when young.[7] So with probability $2H(s) - 1$, the decision to buy from

[6] Provided that s is sufficiently small or $h(d)$ is uniform and $\frac{s}{D} < \frac{3}{5}$, there exists a unique $\tilde{y}$ which is monotonically increasing in $p_{B,y}$ and decreasing in $p_{A,y}$.

[7] In other words, consumers with $d^2 < -s$ lose out whilst those with $d^2 > s$ gain. But consumers do not know ex ante what their d^2 will be.

A when young causes a consumer to pay $p_{A,o}$ instead of $p_{B,o}$ when old. And we know from Lemma 21 that a small increase in $\tilde{y}$ raises $p_{A,o} - p_{B,o}$ by an amount $2\frac{h(0)[2H(s)-1]}{3h(s)}$. So a young consumer who responds to a small cut in $p_{A,y}$ by buying from A, suffers an expected future loss of $\delta_c \left[2H(s)-1\right] 2\frac{h(0)[2H(s)-1]}{3h(s)}$.

The price paid by young consumers can be derived as follows. Retailer A chooses $p_{A,y}$ to maximise $p_{A,y}H\left(\tilde{y}\right) + \delta_f \Pi_{A,o}$ - the sum of expected profits earned from a given cohort of young consumers. Taking the relevant first order conditions and solving for a symmetric equilibrium gives the following[8]:

Lemma 23 *Young consumers pay a price*

$$p_{A,y} = p_{B,y} = \frac{1}{2h\left(0\right)} + \underbrace{\frac{\left[2H(s)-1\right]}{3h(s)}\left[\delta_c[2H(s)-1] - \delta_f\right]}_{\textit{Total lock-in effect}}$$

Young consumers pay the standard (zero-switching cost) Hotelling price plus a 'lock-in' term. Switching costs affect this term in two ways:

1. **Consumer influence** $\frac{[2H(s)-1]}{3h(s)}\delta_c[2H(s)-1]$, which *raises* price

2. **Retailer influence** $-\frac{[2H(s)-1]}{3h(s)}\delta_f$, which *lowers* price

Consumers fear being locked into a high-price retailer in the second-period, and therefore respond less to price changes when young. Recall that demand sensitivity is reduced

[8]Profit is quasiconcave in $p_{A,y}$ if s is not too large and/or $h\left(d\right)$ is close to being uniform.

by the probability of lock-in $2H(s)-1$, multiplied by the difference in price $2\frac{\partial p_{A,o}}{\partial \tilde{y}}$ caused by a small change in first-period market shares.

Retailers know that first-period market share is valuable. In particular a small increase in market share among young consumers is worth $\delta_f \frac{\partial \Pi_{A,o}}{\partial \tilde{y}} = \delta_f \frac{\partial p_{A,o}}{\partial \tilde{y}}$, so firms compete first-period prices down by a proportional amount.

4.4.2 What is the Net Effect of Switching Costs?

The model has three interesting conclusions. Firstly, small switching costs are unambiguously pro-competitive. Secondly, prices may be lower in both periods and follow a pattern of 'ripoffs-then-bargains'. And thirdly, even quite large switching costs can benefit consumers and harm producers.

Small switching costs are pro-competitive and good for consumers. In particular starting from zero, a small increase in the switching cost leaves second-period prices unchanged but strictly reduces first-period prices. Intuitively, marginal old consumers have $d^2 \approx 0$ and their mass is unaffected by a small change in the switching cost (since the density of consumer preferences is symmetric). Retailers cannot exploit old consumers so, using the terminology in Section 4.3.1, the harvesting effect is zero. However small switching costs do affect competition for young consumers. We argued that market share is valuable, so firms compete for young consumers by reducing their price - and this investment effect is first-order. Meanwhile the elasticity effect is only second-order, because the probability of lock-in is essentially zero.

Using Lemma 23 and starting from zero, a small increase in the switching cost therefore transfers an amount $\frac{2\delta_f}{3}$ from retailers to consumers. Since half of old consumers switch supplier, their expected surplus is reduced by an amount $\frac{\delta_c}{2}$. Therefore unless consumers are significantly more patient than firms, $\frac{2\delta_f}{3} > \frac{\delta_c}{2}$ and small switching costs raise consumer surplus.

When the switching cost is larger, all prices may be lower and be characterised by 'ripoffs-then-bargains'. Lemma 23 demonstrates that switching costs always reduce first-period prices unless consumers are particularly patient. Intuitively consumers are rarely locked to their initial supplier, so they care less than firms do about changes in future prices. Equation (4.3) shows that switching costs can also easily reduce second-period prices. Higher switching costs force retailers to compete for consumers with more extreme preferences. If $h(d)$ is U-shaped, consumers with such extreme preferences are more numerous and therefore competition is fiercer. In particular the *harvesting effect* need not lead to higher prices. Moreover depending upon discount factors and preferences, old consumers may pay *lower prices* than young consumers.[9] Figure 4-1 illustrates this with a simple example where $d \in [-1, 1]$ and $h(d) = \frac{13}{30} + \frac{d^2}{5}$.[10]

Finally switching costs can often benefit consumers. For tractability, take the following example:

[9] Only around $s = 0$ can we be sure that prices follow 'bargains-then-ripoffs'. If δ_c is large relative to δ_f and $h(d)$ has an inverted-U shape, then all prices may be higher; however the young may still pay more than the old.

[10] Note that whilst the distribution is logconcave, I assume that the first order conditions are sufficient to characterise equilibrium.

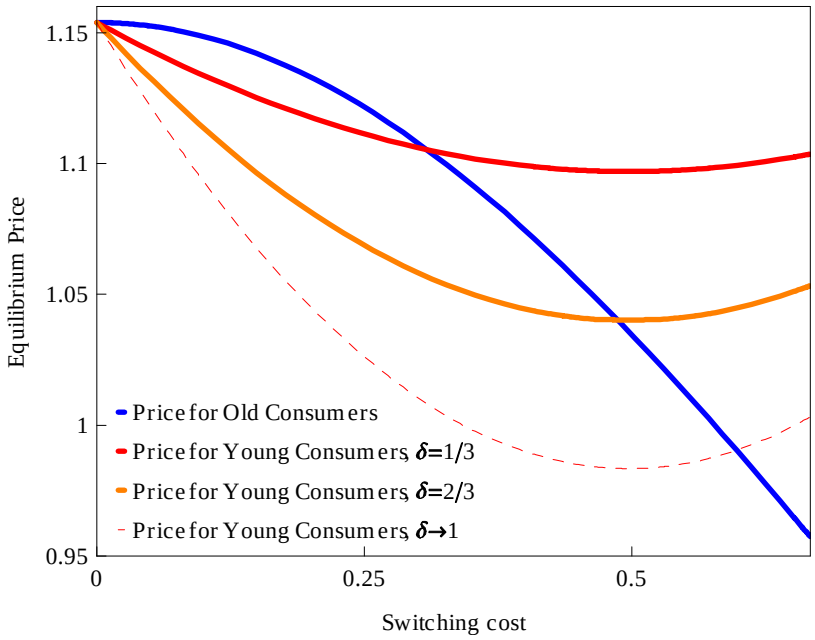


Figure 4-1: 'Ripoffs-then-bargains' Pricing

Example 24 *If consumer preferences are uniform*

- *Average price:* $D - \delta_f \frac{s}{3} + \delta_c \frac{s^2}{3D}$

- *Profit per firm:* $\frac{D}{2}\left[1 + \delta_f\right] + \frac{s}{3}\left[\delta_c \frac{s}{D} - \delta_f\right]$

- *Consumer surplus:* $CS|_{s=0} - \frac{5}{12}\delta_c \frac{s^2}{D} - \delta_c \frac{s}{2} + \delta_f \frac{2}{3} s$ *(where* $CS|_{s=0}$ *is the level of consumer surplus when* $s = 0$*)*

Price and profit are both decreasing in the switching cost whenever $\frac{s}{D} < \frac{1}{2}\frac{\delta_f}{\delta_c}$, and more generally are lower than they would be in the absence of switching costs provided that $\frac{s}{D} < \frac{\delta_f}{\delta_c}$. Intuitively $h(s)$ is constant therefore prices paid by old consumers are unaffected. The forward-looking effect is weaker (and therefore dominated by the firm investment effect) whenever the probability of lock-in is small and/or consumers are impatient. The latter occur when s and δ_c respectively are small. Consumer surplus is

higher than in the absence of switching costs provided $\frac{s}{D} < \frac{1}{5}\left[8\frac{\delta_f}{\delta_c} - 6\right]$, and is strictly increasing in s whenever $\frac{s}{D} < \frac{4}{5}\frac{\delta_f}{\delta_c} - \frac{3}{5}$. Therefore if consumers and firms are equally as patient, there is a wide range of parameters where *relatively small* switching costs increase competition and benefit consumers.

4.5 Full Discrimination

In this Section firms are permitted to use behaviour-based price discrimination. In contrast to models with homogeneous goods, I find that switching costs always hurt retailers and often benefit consumers.

Behaviour-based discrimination simply means that a firm may charge different prices to each of the three groups of consumer. The prices charged by firm i in any time period t are denoted as follows: young consumers without a prior product attachment pay $p_{i,y}$; old 'locked' consumers who previously bought from firm i pay p_l; and old 'unlocked' consumers who previously bought from the rival firm j pay p_u.

4.5.1 Solving for prices

The prices paid by old consumers are particularly easy to characterise. Consider an old consumer who previously bought from firm i. He will stay with firm i provided that $V_i - p_l \geq V_j - p_u - s$ or alternatively $d \leq p_u - p_l + s$. Firm i therefore sets p_l in order to maximise $p_l\left[1 - H\left(p_l - p_u - s\right)\right]$, while firm j sets p_u to maximise $p_u H\left(p_l - p_u - s\right)$.

A unique pair of prices p_l and p_u satisfy the relevant first order conditions:

$$p_l = \frac{1 - H(p_l - p_u - s)}{h(p_l - p_u - s)} \text{ and } p_u = \frac{H(p_l - p_u - s)}{h(p_l - p_u - s)}$$

In contrast to the partial discrimination model, p_l and p_u are the same for both retailers, and independent of how many consumers are locked in to each firm. Since firms can tailor their prices to a consumer's purchase history, there are effectively two different 'aftermarkets' - one for consumers locked into A, and another for consumers locked into B. Intuitively the size of those two markets merely scale up demands, but have no effect on marginal pricing incentives. See for example Taylor [69]. When $s = 0$ consumers pay the standard Hotelling price $\frac{1}{2h(0)}$ regardless of who they purchase from. Otherwise

$$p_u + s > p_l > p_u$$

Switching costs give a retailer some monopoly power over its locked consumers, therefore $p_l > p_u$. But it never benefits a retailer to exploit its customers too much, so $p_u + s > p_l$ and consumers tend to stick with their original supplier.[11] Moreover consumers only switch supplier to get a better product - not to get a better price. This obviously contrasts with models like Chen [16] in which products are undifferentiated.

Lemma 25 *When the switching cost increases:*

[11] Intuitively the retailer's demand curve on its locked consumers is shifted outwards (compared with $s = 0$). The retailer optimally shares this higher demand between a higher price and a higher quantity.

- *a firm charges its locked-in consumers* more
- *but charges other people* less
- *consumers become* more *likely to buy from their original supplier*

A higher switching cost makes it easier for a firm to exploit its own customers but harder to poach business from a rival. Therefore p_l increases whilst p_u decreases. Since $p_l - p_u - s$ is decreasing in s, the equilibrium also has the intuitive feature that fewer consumers switch when the cost of doing so increases. A retailer earns profit $p_l [1 - H(p_l - p_u - s)]$ on its own locked consumers, and profit $p_u H(p_l - p_u - s)$ on those who are locked to the rival. It follows from Lemma 25 that the former is increasing in s whilst the latter is decreasing; this is an important difference with some other papers.[12] Since the future value of locking in a young consumer is simply $p_l [1 - H(p_l - p_u - s)] - p_u H(p_l - p_u - s)$, it follows that

Corollary 26 *The future value of a locked-in young consumer is strictly increasing in the switching cost*

The prices paid by young consumers are also simple to derive. Since the expected payoff (when old) of being locked into A or B is the same, a young consumer born in period t chooses to buy from the retailer offering the highest period-t payoff.[13]

[12] With homogeneous products, Bouckaert *et al* [13] show that when everybody's switching cost is scaled up (as in Chen [16] and Gehrig and Stenbacka [31]), p_l *increases* and the proportion of switchers *stays constant*. But if switching costs are increased by a lump-sum, comparative statics are as in the current model.

[13] Note that future preferences are not forecastable, and whichever firm the consumer locks into, he faces the same prices p_l and p_u.

Consequently a young consumer buys from firm A if $d^t \leq p_{B,y} - p_{A,y} = \tilde{y}$, and buys from firm B otherwise. In each period, firm A therefore chooses $p_{A,y}$ to maximise the discounted total profits from any given cohort of young consumers:

$$H(\tilde{y})\left\{p_{A,y} + \delta_f p_l\left[1 - H(p_l - p_u - s)\right]\right\} + \left[1 - H(\tilde{y})\right]\delta_f p_u\left[H(p_l - p_u - s)\right]$$

Discounted profits come from two sources. A mass $H(\tilde{y})$ of young consumers buy from firm A. They pay $p_{A,y}$ when young, and (as explained above) pay firm A an expected amount $p_l\left[1 - H(p_l - p_u - s)\right]$ when old. A mass $1 - H(\tilde{y})$ of young consumers buy from A's rival, but (due to switching) pay firm A an expected amount $p_u H(p_l - p_u - s)$ when old. One can show that there is a unique equilibrium price $p_{A,y} = p_{B,y} = p_y$ as follows

$$p_y = \frac{1}{2h(0)} - \delta_f \underbrace{\left[p_l\left[1 - H(p_l - p_u - s)\right] - p_u\left[H(p_l - p_u - s)\right]\right]}_{\text{Future value of locking in a young consumer}} \tag{4.5}$$

Notice that young consumers pay the standard Hotelling (zero switching cost) price of $\frac{1}{2h(0)}$, minus a discount - where the discount is exactly equal to the discounted future value of a locked-in consumer.

Remark 27 *As the switching cost increases, the price paid by young consumers* falls

A higher switching cost makes it easier for a firm to exploit locked-in consumers, but more difficult to sell to consumers who previously bought from its rival. This encourages retailers to compete more aggressively for young consumers, which causes p_y to fall.

4.5.2 What is the Net Effect of Switching Costs?

The model has three interesting conclusions. Firstly, small switching costs are unambiguously pro-competitive. Secondly, profits are monotonically decreasing in the size of the switching cost. And thirdly even large switching costs can benefit consumers and harm producers.

Lemma 28 *Start from $s = 0$ and slightly increase the switching cost*

- *The average price paid by old consumers is unchanged*
- *But the price paid by young consumers falls by $\frac{2\delta_f}{3}$*

Small switching costs are pro-competitive and (probably) benefit consumers. It is simple to show that p_l increases by $\frac{1}{3}$ whilst p_u decreases by $\frac{1}{3}$. Intuitively marginal consumers have $d^2 \approx 0$ and therefore small switching costs do not change second-period thickness of demand. Instead switching costs simply shift demand towards a consumer's incumbent supplier. Since optimum prices are just the ratio of demand to demand thickness, it follows that p_l and p_u are shifted by *equal amounts but in opposite directions.* The average second-period price is therefore unchanged, since half of old consumers switch supplier whilst the other half don't; the harvesting effect cancels out.[14] Since p_l and p_u are independent of initial market shares, the elasticity effect also has no influence on prices. Therefore only the investment effect is relevant: profits on a locked consumer rise

[14] Cabral [15] makes a similar point in a model with infinitely-lived consumers. He analyses long-run behaviour when $s \approx 0$ or $\delta_f \approx 1$. He does not have an initial stage where firms compete to lock consumers in, and so does not study lifetime prices or profits.

by $\frac{1}{3}$ whilst profits on an unlocked consumer fall by $\frac{1}{3}$. An additional young consumer is therefore worth $\delta_f \frac{2}{3}$, and first-period prices are competed down by this amount. On average small switching costs affect prices in the same way as under partial discrimination - so profits fall and consumer surplus usually increases.

Retailers are always hurt by higher switching costs - and this is true regardless of discount factors, the shape of consumer preferences, or the size of s. The (per firm) discounted profits earnt on each cohort of consumers can be shown to equal

$$\frac{1}{4h\left(0\right)} + \underbrace{\delta_f \ p_u \left[H\left(p_l - p_u - s\right)\right]}_{\text{Profit on unlocked consumers}} \tag{4.6}$$

We can interpret (4.6) as follows: a firm earns $p_u \left[H\left(p_l - p_u - s\right)\right]$ per old consumer, plus an additional rent $p_l \left[1 - H\left(p_l - p_u - s\right)\right] - p_u \left[H\left(p_l - p_u - s\right)\right]$ if the consumer is locked-in to it. However this rent is fully competed away in the first period, so the true value of old consumers is simply their discounted expected earnings if they are attached to the rival - namely $\delta_f \ p_u \left[H\left(p_l - p_u - s\right)\right]$. We know that the latter - and therefore total discounted profits - is monotonically decreasing in the size of the switching cost. Intuitively when the switching cost is higher, it is harder to make money by poaching customers from a rival.

When firms sell *undifferentiated* products and practise behaviour-based discrimination, switching costs are unambiguously good for firms (see for example Bouckaert *et al* [13]); the same is not true when products are differentiated. Regardless of product differentiation, total profits equal the sum of a). standard competitive profits and b). the

discounted value of a consumer locked to the rival (which is usually strictly positive). It follows that when products are homogeneous, switching costs differentiate retailers and enable them to do better than Bertrand competition (which occurs when $s = 0$). The comparison seems ambiguous when products are differentiated: but since the switching costs make it harder to poach customers from a rival, retailers actually lose out.

Switching costs often reduce average prices. It is clear from (4.5) that young consumers always pay less when the switching cost is larger. The unweighted average price paid by old consumers is $\frac{p_l+p_u}{2} = \frac{1}{2h(p_l-p_u-s)}$ and just as with partial discrimination, this is decreasing whenever consumer preferences are U-shaped.[15] By contrast the weighted average (transaction) price paid by old consumers is

$$p_l \left[1 - H\left(p_l - p_u - s\right)\right] + p_u \left[H\left(p_l - p_u - s\right)\right] \tag{4.7}$$

and this is probably increasing in the switching cost.[16] This is because when the switching cost increases, incumbents charge a higher price and take a larger share of consumer demand. Finally the overall transaction price in a given period is equal to

$$\frac{1}{4h\left(0\right)} + \frac{p_l \left[1 - H\left(p_l - p_u - s\right)\right]\left[1 - \delta_f\right]}{2} + \frac{p_u \left[H\left(p_l - p_u - s\right)\right]\left[1 + \delta_f\right]}{2} \tag{4.8}$$

When $\delta_f = 1$, (4.8) reduces to (4.6) and average price is monotonically decreasing in the switching cost. At the other extreme, when $\delta_f = 0$, (4.8) is almost the same as (4.7)

[15] Recall that $p_l - p_u - s$ is strictly decreasing in s.

[16] It definitely increases in s whenever $h\left(d\right)$ is uniform or inverted-U.

and is therefore probably increasing in the switching cost. More generally switching costs reduce average prices whenever the *investment effect* is sufficiently strong - and this happens when δ_f is large enough. Inspection of (4.8) suggests that switching costs are more likely to result in lower prices when preferences are U-shaped. Firstly because when $s = 0$, marginal consumers are scarce so competition is weak and prices high. Secondly because when $s > 0$, marginal old consumers have extreme preferences, and these preferences are relatively common. So competition is fiercer and the average price paid by an old consumer cannot rise too much.[17]

Example 29 *Suppose preferences are uniformly distributed*

- Average price: $D - \delta_f \frac{s}{3} + \frac{s^2}{18D}$

- Profit per firm: $\frac{D}{2}\left[1 + \delta_f\right] + \delta_f \frac{s}{3}\left[\frac{s}{6D} - 1\right]$

- Consumer surplus: $CS|_{s=0} + \delta_f \frac{2s}{3} + \delta_c \frac{s^2}{36D} - \delta_c \frac{s}{2}$ *(where* $CS|_{s=0}$ *is the level of consumer surplus when* $s = 0$*)*

Average price is decreasing in s provided that $\frac{s}{D} < 3\delta_f$, and is lower than without any switching cost provided that $\frac{s}{D} < 6\delta_f$. Hence switching costs are pro-competitive for a wide range of parameters. Similarly consumer surplus is increasing in s ($s \neq 0$) provided that $\frac{s}{D} > 9 - 12\frac{\delta_f}{\delta_c}$ and is higher than without switching costs whenever $\frac{s}{D} > 18 - 24\frac{\delta_f}{\delta_c}$. Surprisingly large switching costs are more beneficial to consumers. Intuitively, a large switching cost means that locked consumers are more valuable, and leads to large price

[17] Of course this reasoning is only suggestive, since the shape of $h(d)$ affects $H(d)$ and with it the equilibrium prices.

reductions for young consumers. Whilst old consumers pay on average more, an old consumer receives a more favourable price when he switches supplier. This makes a higher switching cost less painful for consumers. Note that if $\delta_c \leq \delta_f$ then switching costs are always beneficial to consumers - the driving force seems to be that firms compete so fiercely for young consumers that they end up transferring rents to consumers.

4.6 No Discrimination

In this section I analyse the effect of switching costs when price discrimination is either illegal or otherwise infeasible. The notation p_i^t is used to denote the (single) price charged in time period t by firm i ($i = A, B$). Unlike in the partial and full discrimination models, a small change in p_i^t will affect prices and profits far into the future. This complicates the analysis, and I therefore simplify matters by assuming that d (the difference in consumer valuations) is uniformly distributed on the interval $[-D, D]$. I also assume that the switching cost parameter is no more than half of the maximum possible product differentiation - in other words $\frac{s}{D} = \tilde{s} \leq \frac{1}{2}$.[18]

In solving the model, I look for a symmetric Markov Perfect Equilibrium (MPE) in which a firm's price p_i^t is a linear function of the state variable. Beggs and Klemperer [10] for example use this approach. It gives a tractable equilibrium which is readily comparable to those found in the full and partial discrimination models. The main

[18] This bound can be increased to (almost) $\sqrt{\frac{1}{3}}$ without too much extra effort. Note also that the partial discrimination model with uniform preferences required $\tilde{s} < \frac{3}{5} \approx \sqrt{\frac{1}{3}}$.

disadvantage is that, at least in principle, there could be other (non-linear) MPE and I do not account for these. The natural state variable in period t is the number of old consumers who previously bought from firm A. Using the standard notation, suppose that young consumers at time $t-1$ bought from firm A if $d_y^{t-1} \leq \tilde{y}^{t-1}$ and bought from firm B otherwise. Then prices (policy functions) satisfy the following equations:

$$p_A^t = J + K\tilde{y}^{t-1} \tag{4.9}$$

$$p_B^t = J - K\tilde{y}^{t-1} \tag{4.10}$$

It is well-known (also proved in the appendix) that when policy functions are linear, value functions are quadratic. Here, a firm's valuation is just the discounted infinite sum of its profits. The value functions are therefore posited to have the following form:

$$V_A^t = M + N\tilde{y}^{t-1} + R\left(\tilde{y}^{t-1}\right)^2 \tag{4.11}$$

$$V_B^t = M - N\tilde{y}^{t-1} + R\left(\tilde{y}^{t-1}\right)^2 \tag{4.12}$$

The set-up differs from Beggs and Klemperer [10] since they assume the switching cost to be so large that no consumer ever changes supplier. It also differs from Doganoglu [21] and Viard [76] since they only solve around $s = 0$ and/or numerically, whereas I solve analytically for general levels of switching costs, and obtain stronger results which also hold out of steady state.

4.6.1 Analysis

I look for an equilibrium in which old consumers always switch with positive probability. I then use the method of undetermined coefficients to solve for the five unknowns (J, K, M, N, R), and demonstrate that the resulting equilibrium prices have intuitive properties.

In contrast to Beggs and Klemperer [10], I look for an equilibrium in which some old consumers always switch supplier. In particular, wherever a young consumer buys from, there must be a strictly positive probability that they switch when they become old. We know that in period t, some old consumers switch from A to B provided that $p_B^t - p_A^t + s < D$, and some others switch from B to A provided that $-D < p_B^t - p_A^t - s$. In contrast to Farrell and Shapiro [26], I also look for an equilibrium in which young consumers spread across the two firms. We know that period-t young consumers buy from firm A provided that they are located below the threshold

$$\tilde{y}^t = \frac{p_B^t - p_A^t}{1 - \delta_c \left[2H\left(s\right) - 1\right] \left[\frac{\partial p_B^{t+1}}{\partial \tilde{y}^t} - \frac{\partial p_A^{t+1}}{\partial \tilde{y}^t}\right]} = \frac{p_B^t - p_A^t}{1 + 2K\delta_c \tilde{s}} \tag{4.13}$$

where the last equality makes use of (4.9) and (4.10). It is clear that in order for $\tilde{y}^t \in (-D, D)$ we require (and I therefore assume) that $1 + 2K\delta_c\tilde{s} \neq 0$. It follows that we can write firm A's period-t demand $D_A^t\left(p_A^t, p_B^t, \tilde{y}^{t-1}\right)$ as:

$$H\left(\tilde{y}^{t-1}\right) H\left(p_B^t - p_A^t + s\right) + \left[1 - H\left(\tilde{y}^{t-1}\right)\right] H\left(p_B^t - p_A^t - s\right) + H\left(\frac{p_B^t - p_A^t}{1 + 2\delta_c K \frac{s}{D}}\right) \tag{4.14}$$

and we let $\pi_A^t\left(p_A^t, p_B^t, \tilde{y}^{t-1}\right) = p_A^t D_A^t\left(p_A^t, p_B^t, \tilde{y}^{t-1}\right)$ denote firm A's period-t flow profits. The task now is to find five equations which jointly define the unknown variables $\{J, K, M, N, R\}$.

Pricing strategies must be subgame perfect, and imposing this allows us to derive two equations. Assume that firm B follows the strategy $p_B^t = J - K\tilde{y}^{t-1}$. Then in every subgame, firm A should maximise its value function by setting the price $p_A^t = J + K\tilde{y}^{t-1}$. Since firm A's value function is just the sum of flow profits and discounted continuation valuation, subgame perfection requires that

$$J + K\tilde{y}^{t-1} = \arg\max_{p_A^t} \pi_A^t\left(p_A^t, p_B^t, \tilde{y}^{t-1}\right) + \delta_f V_A^{t+1}\left(\tilde{y}^t\right) \tag{4.15}$$

Maximising (4.15) with respect to p_A^t gives a first order condition. We assumed that firms would optimally use the pricing strategies $p_A^t = J + K\tilde{y}^{t-1}$ and $p_B^t = J - K\tilde{y}^{t-1}$. Substituting these behaviours into the first order condition gives us an equation $\alpha_1 + \alpha_2\tilde{y}^{t-1} = 0$, where α_1 and α_2 are functions of δ_f, δ_c and the five unknowns. Since $\alpha_1 + \alpha_2\tilde{y}^{t-1} = 0$ has to hold for every $\tilde{y}^{t-1}$, it must be the case that $\alpha_1 = \alpha_2 = 0$. Imposing this gives us two equations - listed as (4.25) and (4.26) in the appendix. We can also show that A's profit function is quasiconcave in price provided that $\tilde{s} \leq \frac{1}{2}$.

We can then derive three more equations by combining information from the pricing strategies and value function. In particular we know that since $\alpha_1 + \alpha_2\tilde{y}^{t-1} = 0$, then firm A optimally follows the pricing strategy $p_A^t = J + K\tilde{y}^{t-1}$ and (by symmetry) firm B optimally uses the pricing strategy $p_B^t = J - K\tilde{y}^{t-1}$. By assumption we can also write

A's value in period $t+1$ as $V_A^{t+1} = M + N\tilde{y}^t + R\left(\tilde{y}^t\right)^2$. This means that A's period-t value function can be written as:

$$\pi_A^t\left(p_A^t, p_B^t, \tilde{y}^{t-1}\right) + \delta_f V_A^{t+1}\left(\tilde{y}^t\right)\Big|_{\substack{p_A^t = J + K\tilde{y}^{t-1},\, V_A^{t+1} = M + N\tilde{y}^t + R\left(\tilde{y}^t\right)^2 \\ p_B^t = J - K\tilde{y}^{t-1}}} \tag{4.16}$$

which, after simplifications, can be expressed in the form $\alpha_3 + \alpha_4\tilde{y}^{t-1} + \alpha_5\left(\tilde{y}^{t-1}\right)^2$. Therefore the posited value function $V_A^t = M + N\tilde{y}^{t-1} + R\left(\tilde{y}^{t-1}\right)^2$ in (4.11) can only be correct for every $\tilde{y}^{t-1}$ if in fact $\alpha_3 = M$, $\alpha_4 = N$ and $\alpha_5 = R$. Imposing these three conditions gives us equations (4.27), (4.28) and (4.29) in the appendix.

To summarise, subgame perfection and profit maximisation jointly allow us to derive five equations in five unknowns. Since these equations are lengthy, all manipulations are relegated to the appendix. Two equations involve only R and K (namely $\alpha_2 = 0$ and $\alpha_5 = 0$); we can easily eliminate R from them and show that K must satisfy the following equation

$$\frac{\tilde{s}}{2}\left[1 + 2\delta_c K\tilde{s}\right]^3 - 3K\left[1 + \delta_c K\tilde{s}\right]\left[1 + 2\delta_c K\tilde{s}\right]^2 + 4\delta_f K^3\left[1 + \delta_c K\tilde{s}\right] = 0 \tag{4.17}$$

This is quartic and has four real roots, which lie in the intervals $\left(-\infty, -\frac{1}{\delta_c\tilde{s}}\right)$, $\left(-\frac{1}{\delta_c\tilde{s}}, -\frac{1-\tilde{s}}{2}\right)$, $\left(\frac{\tilde{s}}{6}, \frac{\tilde{s}}{5}\right)$ and $\left(\frac{1}{2}, \infty\right)$. However when constructing the demand function (4.14), I assumed that old consumers switch with positive probability, and that young consumers spread between the two firms. Three of the roots are incompatible with these assumptions.[19]

[19] Old consumers only switch with positive probability if $K < \left|\frac{1-\tilde{s}}{2}\right|$. To see this, notice that assuming

Therefore there is a unique value for K, namely the root of (4.17) which lies in the interval $\left(\frac{\tilde{s}}{6}, \frac{\tilde{s}}{5}\right)$. **Indeed the whole system of equations has a unique solution**. This is because the other four unknown coefficients (J, M, N, R) can be expressed solely as functions of K, and in particular as follows:

$$\frac{J}{D} = \frac{1 + 2\delta_c K\tilde{s} + \delta_f K}{1 + \delta_c K\tilde{s} + \frac{\tilde{s}}{2}\delta_f} \tag{4.18}$$

$$M = \frac{J}{1-\delta} \tag{4.19}$$

$$N = \frac{\frac{\tilde{s}}{2}\left[1 + 2\delta_c K\tilde{s}\right] - K\left[1 + \delta_c K\tilde{s}\right]}{1 + \delta_c K\tilde{s} + \frac{\tilde{s}}{2}\delta_f} \tag{4.20}$$

$$R = \frac{K^2}{D}\frac{1 + \delta_c K\tilde{s}}{1 + 2\delta_c K\tilde{s}} \tag{4.21}$$

Unfortunately it is hard to interpret equations (4.18) - (4.21) except in the (uninteresting) case of $s = 0$ (when as expected $K = N = R = 0$ and $J = D$). However rewriting (4.18) as $\frac{J}{D} = 1 + \frac{\delta_c K\tilde{s} - \delta_f N}{1 + \delta_c K\tilde{s}}$, it is clear that steady state price is driven by the usual investment and elasticity effects. Therefore when N is larger, a small increase in market share is more valuable so firm competition drives the price down. And when $K\tilde{s}$ is larger, young consumers understand that (1) a low-price firm will try to exploit them more heavily in the future, and (2) they are less likely to switch away and avoid this

the firms follow (4.9) and (4.10), the switching condition $p_B^t - p_A^t + s < D$ can be rewritten as $-2K\tilde{y}^{t-1} + s < D$. This holds for all $\tilde{y}^{t-1}$ provided that i). $K > 0$ and $2KD + s < D$ or ii). $K < 0$ and $2KD > s - D$. Both are equivalent to $K < \left|\frac{1-\tilde{s}}{2}\right|$; we can do the same for the switching condition $-D < p_B^t - p_A^t - s$. So if a K exists, it must be the root of (4.17) that lies in $\left(\frac{\tilde{s}}{6}, \frac{\tilde{s}}{5}\right)$. We therefore need to check that young consumers spread between the two firms when $K \in \left(\frac{\tilde{s}}{6}, \frac{\tilde{s}}{5}\right)$. They do provided that $\frac{p_B^t - p_A^t}{1 + 2K\delta_c\tilde{s}} = \frac{-2K\tilde{y}^{t-1}}{1 + 2K\delta_c\tilde{s}} \in (-D, D)$, which clearly holds.

exploitation - so their demand becomes less elastic. Meanwhile firms earn per-period profit of J in steady state, so M is simply the discounted sum of these profits.

Summary 30 *There exists an equilibrium with the following properties*

- *A firm with a larger installed base charges a higher price*
- *Installed base is valuable, and increasingly so*
- *Each firm charges a positive price and has a positive value*
- *A's equilibrium market share, sh_A^t, satisfies the equation*

$$sh_A^t - \frac{1}{2} = -\frac{2K}{1+2\delta_c K\tilde{s}}\left[sh_A^{t-1} - \frac{1}{2}\right] \tag{4.22}$$

and therefore converges (non-monotonically) to $\frac{1}{2}$

Since $K > 0$, a firm's optimal price in period $t+1$ is positively related to its share of young consumers in period t. The intuition again follows from the usual tradeoff faced by a firm: price high and exploit current customers, or price low and attract people who can be exploited later on. More interestingly, notice that in this model $K \in \left(\frac{\tilde{s}}{6}, \frac{\tilde{s}}{5}\right)$, whereas in the partial discrimination model $K = \frac{\tilde{s}}{6}$.[20] Therefore pricing incentives in this more complicated model are very close to those in the simpler one. Similarly we can show that $N > 2RD$, and this implies that a firm's period-t value function is increasing

[20] Precisely, $\frac{\partial\left(p_{A,o}^t + p_{A,y}^t\right)}{\partial \tilde{y}^{t-1}} = \frac{\tilde{s}}{6}$ where $p_{A,o}^t + p_{A,y}^t$ is firm A's average period-t price.

in its share $\tilde{y}^{t-1}$ of young consumers in the previous period. Convexity of the value function is also common - see for example Beggs and Klemperer [10].

Switching costs have relatively small effects on optimal prices. Although the switching cost can reach as much as one half of total product differentiation, I find that the (normalised) steady state price $\frac{J}{D} \in \left[\frac{13}{15}, \frac{22}{21}\right]$ (compared to $\frac{J}{D}\big|_{s=0} = 1$). In particular prices are well above marginal cost, and since $K < \frac{\tilde{s}}{5} < \frac{1}{10}$, even a new entrant with no installed base will never set a (normalised) price below $0 \cdot 7\dot{6}$. Moreover the gap between the highest and lowest (normalised) prices in the market can never exceed $2K < \frac{1}{5}$. Consequently even a new entrant charges a price that is at least 80% of that charged by an established retailer. Unsurprisingly, therefore, firm values are positive and significant.

The market converges to a situation where the two firms split demand equally and charge the same price J. To understand this, first use (4.9), (4.10) and (4.13) to write

$$\tilde{y}^t = -\frac{2K}{1 + 2K\delta_c \tilde{s}} \tilde{y}^{t-1}$$

It is clear that in steady state $\tilde{y} = 0$, and using equations (4.9) and (4.10), each firm charges a price J. Moreover since $-\frac{2K}{1+2K\delta_c\tilde{s}} \in (-1, 0)$, the market always converges to this steady state, albeit non-monotonically. Non-monotonicity is explained as follows: if $\tilde{y}^{t-1} > 0$ then $p_A^t > p_B^t$ and period-t young consumers respond by substituting towards firm B - resulting in $\tilde{y}^t < 0$. Convergence towards $\tilde{y} = 0$ is the result of small switching costs and changing tastes: when $\tilde{y}^{t-1} > 0$, firm A cannot raise its period-t price by too much or it will lose a lot of customers. Therefore although $p_B^t - p_A^t$ is a linear function

of $\tilde{y}^{t-1}$, it is not very responsive to changes in $\tilde{y}^{t-1}$. This implies that period-t young consumers see less difference between the two retailers than their predecessors did, and therefore spread more evenly between the two firms. Consequently $|\tilde{y}^t| < |\tilde{y}^{t-1}|$, and a firm's share of *young* consumers converges non-monotonically to steady state.

Total market share also converges non-monotonically to steady state. Mathematically this is because in equation (4.22), $-\frac{2K}{1+2\delta_c K\tilde{s}} \in (-1, 0)$. Intuitively, the switching cost makes it relatively difficult to change suppliers, so there is a positive relationship between a firm's share of young consumers in period $t-1$, and its *total* market share in the following period t. Since the former (share of young consumers at $t-1$) behaves non-monotonically, so does the latter (total market share at t). However the firm with the larger share of $t-1$ young consumers will exploit this in period t by charging a high price, and this naturally depresses its period-t demand. Consequently a firm's period-t *total* market share is always closer to $\frac{1}{2}$ than its share of young consumers in the previous period. Since shares of *young* consumers converge to $\frac{1}{2}$, so does *total* market share.

Lemma 31 *All prices are*

- *decreasing in* δ_f
- *increasing in* δ_c
- *decreasing in* δ *(where* $\delta = \delta_c = \delta_f$*)*

These comparative statics are valid both in and out of steady state, and whilst the proofs are messy, the intuition is very clear. When δ_f is larger, firms care more about

future profits and so the investment effect is stronger; when δ_c is larger, consumers care more about the future repercussions of purchasing from a cheap supplier, so demand becomes less elastic. However as explained previously, consumers are less concerned about future price movements than are firms. Therefore ceteris paribus the investment effect dominates, and prices are decreasing in δ.

4.6.2 Switching Costs Reduce Prices

I begin by showing that switching costs reduce steady state prices under very mild conditions; I also demonstrate that outside of steady state, average prices are often lower as well. I then demonstrate that with further restrictions, even an incumbent monopolist (facing a brand new entrant) charges a lower price when its customers have switching costs.

Steady State Prices

Recalling the discussion on page 126, the marginal period-t young consumer satisfies $\tilde{y}^t = -\frac{2K}{1+2K\delta_c\tilde{s}}\tilde{y}^{t-1}$. The market converges to a steady state in which $\tilde{y} = 0$, and (using the equilibrium pricing equations (4.9) and (4.10)), each firm charges a price of J. It is therefore natural to begin by examining how switching costs affect this steady state equilibrium price.

Lemma 32 *The steady state price J*

- *is decreasing in s around $s = 0$, regardless of δ_c or δ_f*

- *is monotonically decreasing in s provided that* $\delta_c = \delta_f$
- *is always below D provided that* $\tilde{s} \leq \frac{3\delta_f}{2\delta_c}$

Small switching costs always reduce steady state price because (using expressions (4.18) and (4.17)) we can show that $\frac{\partial J}{\partial s}\big|_{s=0} = -\frac{\delta_f}{3}$.[21] Intuitively the harvesting effect is zero - retailers split demand equally and still compete for old consumers with $d^2 \approx 0$. The mass of these marginal old consumers is unchanged, and therefore so are firms' marginal pricing incentives amongst old consumers. As usual the consumer elasticity effect is zero, so only the investment effect operates - and using (4.20) and (4.17) we can show that $\frac{\partial N}{\partial s}\big|_{s=0} = \frac{\delta_f}{3}$. So market share is valuable and retailers reduce price to compete for it.

Large switching costs also reduce steady state price under mild conditions. The harvesting effect is always zero because firms split demand equally and the mass of marginal old consumers is unaffected. Retailers want to attract young consumers and exploit them later; young consumers understand this and are less inclined to buy from cheap suppliers. But only with probability $\tilde{s}$ does a consumer's first-period purchase decision affect their second-period decision, and therefore ceteris paribus the investment effect dominates the elasticity effect. When $\delta_c = \delta_f$ and the switching cost increases, the former effect becomes increasingly dominant and steady state price falls. Moreover the investment effect always dominates provided $\tilde{s} \leq \frac{3\delta_f}{2\delta_c}$. Consequently steady state price

[21] Small switching costs also reduce average price by this amount when firms can partially or fully discriminate - I return to this point later on.

is lower than D (price in the absence of switching costs) even when consumers are three times more patient than retailers.

Prices Outside of Steady State

In this section I demonstrate that all prices may be lower when consumers have switching costs. Of particular interest are the maximum and average prices charged by firms in equilibrium. We know from equations (4.9) and (4.10) that the largest price ever observed in equilibrium is $J + KD$, whilst the average price paid by consumers is equal to $J + K\tilde{y}^{t-1}\left(D_A^t - 1\right)$. Using the expression for D_A^t given in (4.14), this average price can be shown to equal

$$J + \frac{K\left(\tilde{y}^{t-1}\right)^2}{D}\left(\frac{\tilde{s}}{2} - 2K\frac{1+\delta K\tilde{s}}{1+2\delta K\tilde{s}}\right) \tag{4.23}$$

Since equilibrium prices are linear in market share, the average price is largest when $\tilde{y}^{t-1} = \pm D$ and all old customers are locked to one retailer. Throughout this section, 'average price' means (4.23) evaluated at $\tilde{y}^{t-1} = \pm D$.

Lemma 33 *Starting from zero, a small switching cost*

- *always reduces the average price*
- *also reduces the maximum price provided* $\delta_f > \frac{1}{2}$

The harvesting effect matters whenever the market is outside of steady state. In particular the firm with the larger market share exploits its customers and prices above

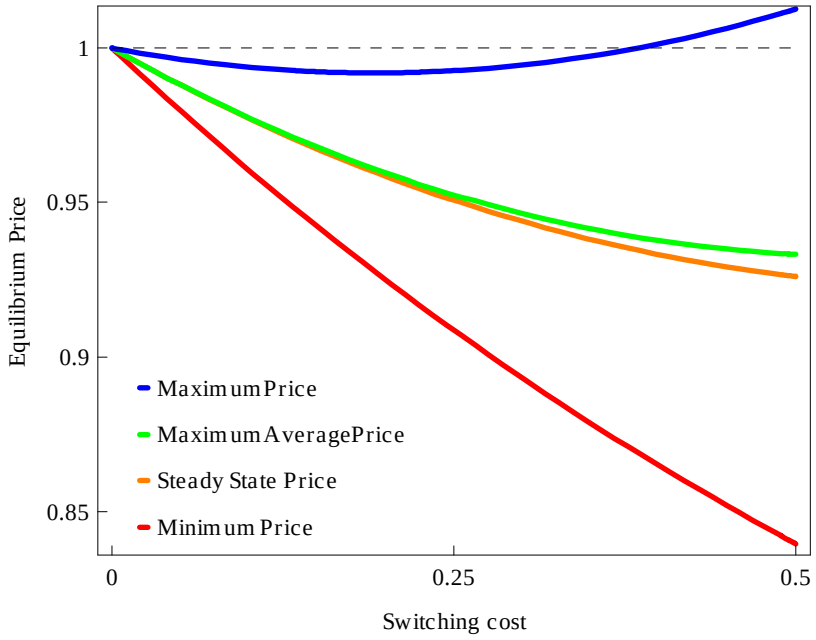


Figure 4-2: Prices under No Discrimination

J whilst the rival prices below J. As the switching cost approaches zero, the two retailers share the market evenly and therefore average price is simply J. By Lemma 32 the average price is therefore strictly decreasing in the switching cost. The maximum price is more ambiguous and depends upon the relative strengths of the harvesting and investment effects. Harvesting increases current profits whilst investing is good for future profits: when retailers are sufficiently patient the investment effect dominates and therefore a small switching cost reduces price. Strikingly, therefore, even an incumbent facing a brand new entrant will optimally reduce its price when consumers have switching costs.

Lemma 34 *Compared with zero switching costs*

- *all prices are lower provided that* $\delta \geq \frac{10}{15(1-\tilde{s})-2\tilde{s}^2}$

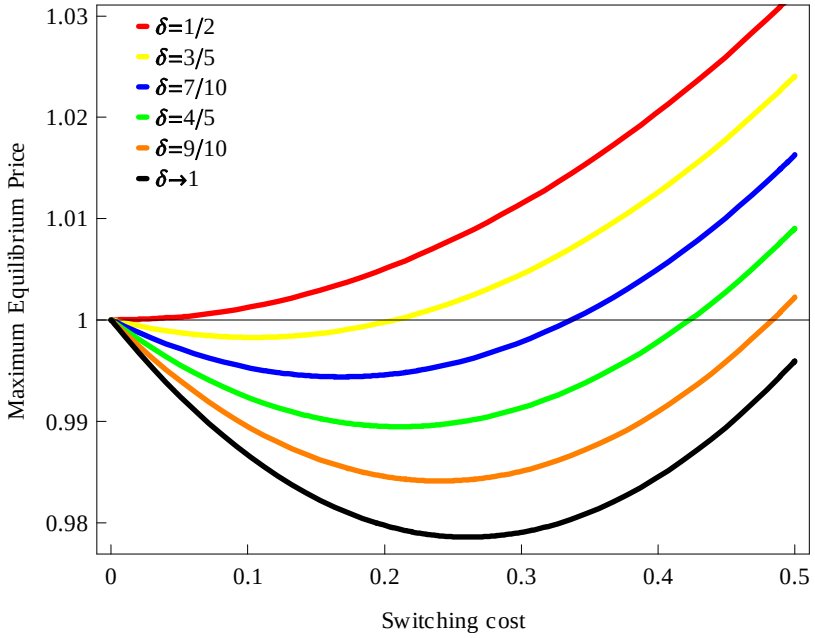


Figure 4-3: Maximum Equilibrium Prices under No Discrimination

- *average price is lower provided that* $\delta \geq \frac{70}{1170-815\tilde{s}-14\tilde{s}^2}$

The harvesting effect pushes prices up whilst the investment and elasticity effects (in aggregate) pull prices down. When discount factors are higher, the latter are collectively stronger and therefore prices are more likely to be lower. Clearly the condition on average price is much weaker, but both conditions are only sufficient and could be substantially weakened if K were solvable analytically.

Figures 4-2 and 4-3 illustrate these points. In Figure 4-2 I plot maximum, minimum, steady state and (maximum) average prices under the assumptions that $\delta_c = \delta_f = \frac{3}{4}$ and $D = 1$. As expected, both minimum and steady state price are monotonically decreasing in the switching cost, and more surprisingly so is the average price. Maximum price follows a U-shape. Intuitively a retailer has less scope to exploit old consumers when

the switching cost is small, so the investment effect dominates and price is decreasing in $\tilde{s}$. However when $\tilde{s}$ is larger, consumers find it harder to switch so the harvesting effect dominates and price increases in $\tilde{s}$. Figure 4-3 focuses exclusively on maximum prices, and again finds this U-shaped relationship. Even for moderately high discount factors like $\delta = \frac{3}{5}$, there is a large range of switching costs such that maximum price is below the level that would prevail when $s = 0$. Moreover, as $\delta \to 1$, an incumbent (former) monopolist charges a lower price even if its old customers have to pay a switching cost equal to half of total product differentiation. These two figures therefore provide two main lessons. Firstly, the expressions in given in Lemma 34 are only sufficient, and can be significantly relaxed when K is solved for explicitly. Secondly, even when switching costs are very large, all market prices can often be lower - a result which contrasts strongly with those derived elsewhere. And of course if consumers are less patient than firms, then prices will be even lower.

4.7 Who Benefits from Price Discrimination?

The three previous sections demonstrated that switching costs are often pro-competitive and beneficial to consumers. An interesting question is whether we can order prices, profits and welfare according to the degree of freedom that firms have over their prices. Whilst complete rankings are often impossible, this section demonstrates that under reasonable conditions, the ability to price discriminate is bad for retailers but good for consumers.

When switching costs are very small, all three types of price discrimination lead to identical outcomes. When retailers can partially or fully discriminate, first-period prices fall by $\frac{2}{3}\delta_f$ and (average) second-period prices are unchanged. If discrimination is impossible, steady state price falls by $\frac{1}{3}\delta_f$ - the average of these two amounts. Intuitively the marginal old consumer has $d^2 = 0$ regardless of how firms price discriminate. Therefore the harvesting effect is always zero, whilst the value of a young consumer (and therefore the investment effect) is the same no matter how much discrimination is possible. This is also true in a model with more than two retailers.[22]

4.7.1 Uniform Preferences

When the switching cost is larger[23] but d is uniform, some progress can be made in ordering prices and comparing welfare under the three different scenarios. Let p_{old}^{full} for example be the average price paid by old consumers when firms can fully discriminate, and let p^{no} be the steady state price when discrimination is infeasible. Then:

Lemma 35 *If* $2\delta_f \geq 3\tilde{s}\delta_c$,

$$p_{young}^{full} < p_{young}^{partial} < p^{no} < p_{old}^{partial} < p_{old}^{full}$$

[22] Take a Spokes Model (Chen and Riordan [18]) with n retailers. Each period a consumer is randomly assigned to a spoke, and then has a preference d distributed symmetrically along the spoke with some density $h(d)$. Then regardless of whether partial or full discrimination is used, small switching costs do not affect the average price paid by old consumers, whilst the price paid by young consumers falls by $\frac{2\delta_f}{2n-1}$.

[23] When $s > 0$, the three effects no longer coincide. The harvesting effect is positive under full discrimination but zero under partial and no discrimination, whilst the opposite is true of the elasticity effect. Profits amongst old consumers behave differently under different pricing regimes, and therefore so does the investment effect.

Since preferences are uniform, prices follow a pattern of 'bargains-then-ripoffs'. When retailers can better price discriminate, they are able to charge a higher (average) price to old consumers. This leads to fiercer competition for young consumers, and therefore the sizes of both the 'bargain' and 'ripoff' components grow with the amount of price discrimination. Explicit solutions can be derived under both partial and full discrimination, so I begin by comparing outcomes in these two cases.

Lemma 36 *Compared to partial discrimination, the ability to fully price discriminate*

- *lowers average price provided* $\delta_c > \frac{1}{6}$
- *lowers firm profits provided* $\frac{\delta_c}{\delta_f} > \frac{1}{6}$
- *increases total surplus provided* $\frac{\delta_c}{\delta_f} < \frac{1}{2}$
- always *increases consumer surplus*

If retailers can fully price discriminate, young consumers pay less but old consumers pay more. On balance average price is reduced and firms are made worse off, provided that consumers are sufficiently patient. Intuitively when δ_c increases, the elasticity effect becomes stronger and this pushes up $p_{young}^{partial}$ - but leaves all other prices unchanged. Hence for sufficiently high δ_c, partial discrimination gives more profits and a higher average price.

Surprisingly full price discrimination can be good for welfare. Switching costs are wasteful, and from a social perspective should only be paid when $|d^2| > s$. Switching

behaviour is therefore efficient with partial discrimination, but not with full discrimination - because too many people change supplier in order to access a lower price. But full discrimination transfers future profits to young consumers - and this can be beneficial if consumers are less patient than firms. This latter effect can dominate the first, and full discrimination can therefore increase welfare.[24] Finally consumers always benefit from full discrimination: if δ_c is low then first-period bargains are valued more than second-period ripoffs, and if δ_c is high then partial discrimination would not deliver a strong initial bargain anyway.[25]

It is difficult to make comparisons with no discrimination because K does not have an explicit solution. Nevertheless we can make two relatively general observations:

Lemma 37 *Full discrimination is*

- *best for consumers provided that* $3\delta_f > \tilde{s}\delta_c$
- *worst for retailers provided that* $6\delta_c > \delta_f$

When firms have more freedom over their pricing, they often earn less profits. Although a monopolist always benefits from the ability to price discriminate, the same need not be true for oligopolists due to strategic interaction (Holmes [38]). In the present model, greater pricing freedom allows firms to extract more money from old consumers,

[24] Indeed welfare may even be higher than in the absence of switching costs.

[25] In a model with homogeneous goods, Chen [16] finds that full discrimination reduces profits and may raise consumer surplus. However this is caused by consumers doing better and firms doing worse in the *mature market* - and hence is very different to my model. Total welfare always falls since $\delta_f = \delta_c$.

and it therefore results in a larger 'ripoff'. But this larger ripoff also strengthens competition for young consumers, and so results in a larger 'bargain'. This bargain is larger relative to partial/no discrimination when δ_c is high, whilst future ripoffs are worth less to a firm if δ_f is small. Hence when δ_c is large relative to δ_f, the ability to fully price discriminate naturally hurts retailers. Similarly consumers benefit from full price discrimination provided they are not too patient (and so do not worry too much about being ripped off later on). Therefore there exists a range of parameters - namely $\delta_f \in \left(\frac{\tilde{s}\delta_c}{3}, 6\right)$ - where discrimination is good for consumers but bad for retailers. As usual the inability to solve explicitly for K implies that the condition $3\delta_f > \tilde{s}\delta_c$ could be weakened significantly.[26]

[26] One would also like to find conditions where consumer surplus monotonically increases (and firm profit monotonically decreases) in the amount of pricing freedom. It turns out that if $\frac{\delta_f}{\delta_c} > \frac{3}{2}\tilde{s} + \delta_f$ then consumer surplus is lower under no discrimination compared to partial discrimination (which in turn is lower than under full discrimination). Similarly it turns out that if $\delta_f \in \left(\frac{3}{2}\delta_c, 6\delta_c\right)$ then profits are highest under no discrimination and smallest under full discrimination.

4.8 Conclusion and Policy Discussion

In a model with differentiated goods and relatively small switching costs, I find a strong presumption that switching costs are pro-competitive and possibly also beneficial to consumers. This research therefore adds to a growing body of literature which is challenging the perceived wisdom. Nonetheless numerical examples suggest that even relatively large switching costs have quite small effects on prices. Confirming this, Dubé *et al* [22] find that even when switching costs are 20% of the final purchase price, there is only a reduction in price of between 3% and 6%.[27] I now suggest two policy implications.

Firstly policymakers should be more concerned with search costs and less concerned with switching costs. Switching costs appear to introduce only mild distortions in prices, whilst (as was seen in the previous Chapter) small search costs can do considerably more damage. If a market is characterised by reasonable switching, it suggests that product differentiation is high and/or switching costs are low. In either case the results from this Chapter are likely to apply, and intervention is unnecessary, and possibly undesirable if the aim is to help consumers. Only in markets where little or no search is observed, are the results of earlier papers such as Beggs and Klemperer [10] applicable.

Secondly, to the extent that policymakers do intervene, it should be on the consumer- and not the firm-side. As in all models, I assume that consumers are forward-looking and are fully aware of possibilities to switch. Ensuring that consumers know this should be

[27] Viard [77] finds that number portability reduces prices, and hence argues switching costs are anti-competitive. The reason may be related to the level of switching cost - in particular it is probably easier to switch between orange juices than between phone suppliers. Therefore orange juice is closer to the current theoretical model.

the priority - attempts to implement policies which result in small changes to switching costs should not be. Moreover, complicated pricing structures should not be banned, but instead be made as transparent as possible. This is because greater abilities to price discriminate ultimately, according to the model presented here, are good for consumers.

4.9 Appendix

4.9.1 Proofs for the Partial Discrimination Model

General Proofs

Proof of Lemma 21 Let $T = p_{B,o} - p_{A,o}$ and note its dependence on $\tilde{y}$. The total derivative of the denominator of (4.1) with respect to $\tilde{y}$ is zero, once we substitute in for $\tilde{y} = T = 0$ (which hold in symmetric equilibrium) and note that $h(s) = h(-s)$ and $h'(s) = -h'(-s)$. The denominator itself is $h(s)$. Partially differentiating the numerator of (4.1) with respect to $\tilde{y}$ and T (and simplifying) we get terms $h(0)[2H(s) - 1]$ and $h(s)$ respectively. Therefore

$$\begin{aligned} \frac{\partial p_{A,o}}{\partial \tilde{y}} &= -\frac{\partial p_{B,o}}{\partial \tilde{y}} = \frac{h(0)[2H(s) - 1]}{h(s)} + \frac{\partial T}{\partial \tilde{y}} \\ \Longrightarrow \quad \frac{\partial p_{A,o}}{\partial \tilde{y}} &= -\frac{\partial p_{B,o}}{\partial \tilde{y}} = \frac{h(0)[2H(s) - 1]}{3h(s)} \end{aligned}$$

Using $\frac{\partial \Pi_{A,o}}{\partial \tilde{y}} = \frac{\partial p_{A,o}}{\partial \tilde{y}} D_{A,o} + p_{A,o} \frac{\partial D_{A,o}}{\partial \tilde{y}}$, the facts that $D_{A,o} = \frac{1}{2}$ and $p_{A,o} = \frac{1}{2h(s)}$ in a symmetric equilibrium, and substituting in from above, we get the expression given in the Lemma. ■

Proof of Lemma 22 Take the equation for $\tilde{y}$ given in the text. The derivative of the righthand side with respect to T (after imposing $T = 0$) is $\delta_c [2H(s) - 1] \frac{\partial T}{\partial \tilde{y}}$. Noting that $\frac{\partial T}{\partial \tilde{y}} = -\frac{2h(0)[2H(s)-1]}{3h(s)}$ gives the desired expression. ■

Proof of Lemma 23 Differentiating $p_{A,y} H(\tilde{y}) + \delta_f \Pi_{A,o}$ with respect to $p_{A,y}$ gives a

first order condition

$$H\left(\tilde{y}\right)+p_{A,y}h\left(\tilde{y}\right)\frac{\partial\tilde{y}}{\partial p_{A,y}}+\delta_f\frac{\partial\Pi_{A,o}}{\partial\tilde{y}}\frac{\partial\tilde{y}}{\partial p_{A,y}}=0$$

Substituting in $\tilde{y}=0$, $p_{A,y}=p_{B,y}$, and our previous formulae for $\frac{\partial\tilde{y}}{\partial p_{A,y}}$ and $\frac{\partial\Pi_{A,o}}{\partial\tilde{y}}$ we get (after rearranging) the expression given in the Lemma. ■

Proof of Existence, Uniqueness and Quasiconcavity

Start by assuming that d is uniformly distributed and $\frac{s}{D}<\frac{3}{5}$.

1). Prices paid by old consumers

a). There is a unique interior equilibrium. The unique solution to (4.1) and (4.2) is $p_{A,o}=D+\frac{s\tilde{y}}{3D}$ and $p_{B,o}=D-\frac{s\tilde{y}}{3D}$. Switching in both directions is possible because $p_{B,o}-p_{A,o}+s<D$ and $p_{B,o}-p_{A,o}-s>-D$ both hold for each $\tilde{y}\in[-D,D]$. Now fix $p_{B,o}=D-\frac{s\tilde{y}}{3D}$ and show that A's profit function is quasiconcave. Whenever

$$-D<p_{B,o}-p_{A,o}-s<p_{B,o}-p_{A,o}+s<D \qquad (INT)$$

A's profit function is clearly concave (because $h'\left(d\right)=0$). So A's profit is strictly increasing in p_A for $p_{A,o}\in\left(p_{B,o}+s-D,D+\frac{s\tilde{y}}{3D}\right)$ and strictly decreasing in $p_{A,o}$ for $p_{A,o}\in\left(D+\frac{s\tilde{y}}{3D},p_{B,o}+D-s\right)$. We must now show that A does not wish to deviate to a $p_{A,o}$ outside the interval INT.

Clearly A never picks $p_{A,o}$ such that it gets zero demand (and profit), so we can

rule out $p_{B,o} - p_{A,o} + s \leq -D$ immediately, and when $H(\tilde{y}) = 0$ we can also rule out $p_{B,o} - p_{A,o} - s \leq -D$. We still need to consider $H(\tilde{y}) > 0$ and the interval INT'

$$p_{B,o} - p_{A,o} - s \leq -D < p_{B,o} - p_{A,o} + s \qquad (INT')$$

A's profit is $p_{A,o}H(\tilde{y})H(p_{B,o} - p_{A,o} + s)$ which is clearly concave. It is simple to check that profit is decreasing in $p_{A,o}$ on the edge of the boundary (when $p_{B,o} - p_{A,o} - s = -D$) since $s < \frac{3}{5}D$. It then follows that profit is decreasing in $p_{A,o}$ elsewhere in INT' (where $p_{A,o}$ is higher).

Clearly A never picks $p_{A,o}$ such that every consumer *strictly* prefers to buy from it, so we can rule out $p_{B,o} - p_{A,o} - s > D$ immediately, and when $H(\tilde{y}) = 1$ we can also rule out $p_{B,o} - p_{A,o} + s > D$. We still need to consider $H(\tilde{y}) < 1$ and the interval INT''

$$p_{B,o} - p_{A,o} - s \leq D < p_{B,o} - p_{A,o} + s \qquad (INT'')$$

A's profit is $p_{A,o}\left[H(\tilde{y}) + \left[1 - H(\tilde{y})\right]H(p_{B,o} - p_{A,o} - s)\right]$ which is clearly concave. It is simple to check that profit is increasing in $p_{A,o}$ on the boundary (when $D = p_{B,o} - p_{A,o} + s$) since $s < \frac{3}{5}D$. It then follows that profit is increasing in $p_{A,o}$ elsewhere in INT'' (where $p_{A,o}$ is lower).

So profit is increasing in $p_{A,o}$ when $p_{A,o} < D + \frac{s\tilde{y}}{3D}$ and decreasing otherwise - hence it is quasiconcave.

b). There are no other equilibria. The 'interior' first order conditions (4.1) and (4.2)

are characterised by the fact that a consumer who bought from firm i when young may, with positive probability, buy either i or $j \neq i$ when old. In particular if $\tilde{y} \in (-D, D)$ then $-D < p_{B,o} - p_{A,o} - s$ and $p_{B,o} - p_{A,o} + s < D$ hold; if $\tilde{y} = -D$ then $-D < p_{B,o} - p_{A,o} - s < D$; and if $\tilde{y} = D$ then $-D < p_{B,o} - p_{A,o} + s < D$.

There never exists an equilibrium in which all consumers buy from one firm.[28] So we need only rule out two cases. Firstly, that when $\tilde{y} \in (-D, D)$ there is no equilibrium with $p_{B,o} - p_{A,o} - s \leq -D$ and $-D < p_{B,o} - p_{A,o} + s$. Secondly, that when $\tilde{y} \in (-D, D)$ there is no equilibrium with $p_{B,o} - p_{A,o} - s < D$ and $D \leq p_{B,o} - p_{A,o} + s$.

In the first case, profits for the two firms are $p_{A,o} H(\tilde{y}) H(p_{B,o} - p_{A,o} + s)$ and for firm B are $p_{B,o}[1 - H(\tilde{y}) + H(\tilde{y})[1 - H(p_{B,o} - p_{A,o} + s)]]$. Solving the corresponding first order conditions, we find that $p_{A,o} = D + \frac{s}{3} + \frac{2}{3}\frac{D-\tilde{y}}{D+\tilde{y}}D$ and $p_{B,o} = D - \frac{s}{3} + \frac{4}{3}\frac{D-\tilde{y}}{D+\tilde{y}}D$. But since $s < \frac{3}{5}D$, it is never possible that $p_{B,o} - p_{A,o} - s \leq -D$ so we find a contradiction.

In the second case, profits are $p_{A,o}[H(\tilde{y}) + [1 - H(\tilde{y})] H(p_{B,o} - p_{A,o} - s)]$ and for firm B are $p_{B,o}[1 - H(\tilde{y})][1 - H(p_{B,o} - p_{A,o} - s)]$. Solving the corresponding first order conditions, we find that $p_{A,o} = \frac{D+\tilde{y}}{D-\tilde{y}}\frac{4}{3}D + D - \frac{s}{3}$ and $p_{B,o} = D + \frac{D+\tilde{y}}{D-\tilde{y}}\frac{2}{3}D + \frac{s}{3}$. But since $s < \frac{3}{5}D$, it is never possible that $D \leq p_{B,o} - p_{A,o} + s$ so we find a contradiction.

2). Prices paid by young consumers

We know that $p_{A,o} = D + \frac{s\tilde{y}}{3D}$ therefore $\Pi_{A,o} = \frac{1}{2D}\left[D + \frac{s\tilde{y}}{3D}\right]^2$. Since $p_{B,o} = D - \frac{s\tilde{y}}{3D}$, we solve the equation for $\tilde{y}$ and show that $\tilde{y} = \frac{p_{B,y} - p_{A,y}}{1 + \frac{2s^2\delta_c}{3D^2}}$. Firm A chooses $p_{A,y}$ to maximise

[28] The intuition is as follows. Suppose that everybody bought from firm A. Then A would increase price until the marginal consumer was indifferent about where to buy. However since s is sufficiently small ($s < \frac{3}{5}D$) firm B could profitably reduce its price and steal some demand, ruling out the putative equilibrium.

$p_{A,y}\left[\frac{D+\tilde{y}}{2D}\right]+\delta_f\frac{1}{2D}\left[D+\frac{s\tilde{y}}{3D}\right]^2$. The first derivative with respect to $p_{A,y}$ is proportional to

$$D+\tilde{y}-\frac{p_{A,y}}{1+\frac{2s^2\delta_c}{3D^2}}+2\delta_f\left[D+\frac{s\tilde{y}}{3D}\right]\frac{s}{3D}\left(\frac{-1}{1+\frac{2s^2\delta_c}{3D^2}}\right)$$

and the second derivative with respect to $p_{A,y}$ is proportional to $\delta_f s^2-9D^2-6s^2\delta_c$ which is negative given the assumption that $s<\frac{3}{5}D$. We can also find B's first order condition, and show that both optimisation conditions hold only when $\tilde{y}=0$. Substituting this back in gives $p_{A,y}=p_{B,y}=D+\frac{2s}{3}\left[\frac{s}{D}\delta_c-\delta_f\right]$.

3). Distributions close to being uniform

The proof is lengthy and therefore omitted, but it proceeds by assuming that $|h''(d)|<\epsilon$ where ϵ is a small positive number. For any $\tilde{y}\in[-D,D]$, there is a unique equilibrium $\{p_{A,o},p_{B,o}\}$ close to $\left\{D+\frac{s\tilde{y}}{3D},D-\frac{s\tilde{y}}{3D}\right\}$ with a quasiconcave profit function. There is also a unique $\tilde{y}$ which is close, and behaves in a similar way, to $\frac{p_{B,y}-p_{A,y}}{1+\frac{2s^2\delta_c}{3D^2}}$. Profit functions in the young market can be shown to be concave and to have a unique equilibrium with $\tilde{y}=0$ and price close to $D+\frac{2s}{3}\left[\frac{s}{D}\delta_c-\delta_f\right]$.

4). Small switching cost

This proof is also omitted, although intuitively as s becomes small, the problem becomes close to a standard zero-switching cost model, which we know is well-behaved.

4.9.2 Proof for the Full Discrimination Model

Proof of prices for old consumers Suppose that $s<D$ and note that since $H(\cdot)$ is logconcave, profits are quasiconcave. Let $T=p_l-p_u$ and combine the first order

conditions:

$$T = \frac{1 - H(T-s)}{h(T-s)} - \frac{H(T-s)}{h(T-s)}$$

When $T = D + s$ then $LHS > RHS$ whilst when $T = -D + s$ then $LHS < RHS$. The righthand side also decreases in T so by continuity there exists a unique T and therefore unique p_l and p_u. If $s = 0$ then $T = 0$; the righthand side increases in s and so therefore does T. Hence $p_l > p_u$. But if $T = s$, $LHS > RHS = 0$ so $T < s$ and therefore $p_l < p_u + s$. We know s raises T - could s (weakly) raise $T - s$? No because the lefthand side would rise but the righthand side fall - therefore $T - s$ decreases in s and consumers are more likely to buy from their initial supplier. ■

Proof of prices for young consumers Note that A's first order condition is:

$$H(\tilde{y}) - p_{A,y} h(\tilde{y}) - h(\tilde{y}) \delta_f \left[p_l \left[1 - H(p_l - p_u - s) \right] - p_u \left[H(p_l - p_u - s) \right] \right]$$

and dividing through by $h(\tilde{y})$ gives:

$$\frac{H(\tilde{y})}{h(\tilde{y})} - p_{A,y} - \delta_f \left[p_l \left[1 - H(p_l - p_u - s) \right] - p_u \left[H(p_l - p_u - s) \right] \right]$$

which is falling in $p_{A,y}$ so profit is quasiconcave. B's first order condition is similar, and both can only hold if $\tilde{y} = 0$. ■

Proof of Lemma 28: *Old consumer prices* Differentiating the first order conditions with respect to s and imposing $s = 0$ and $p_l = p_u = \frac{1}{2h(0)}$ we get $\frac{\partial p_l}{\partial s} = -\frac{\partial p_l}{\partial s} + \frac{\partial p_u}{\partial s} + 1$ and $\frac{\partial p_u}{\partial s} = \frac{\partial p_l}{\partial s} - \frac{\partial p_u}{\partial s} - 1$. Combining these gives $\frac{\partial p_l}{\partial s} = -\frac{\partial p_u}{\partial s} = \frac{1}{3}$. Average price is

$p_l\left[1-H\left(p_l-p_u-s\right)\right]+p_uH\left(p_l-p_u-s\right)$ and upon differentiating with respect to s, is readily seen to be constant around $s=0$. *Young consumer prices* Differentiating $H\left(p_l-p_u-s\right)$ with respect to s and imposing $s=0$ gives $h\left(0\right)\frac{-1}{3}$. The value of a locked consumer is

$$p_l\left[1-H\left(p_l-p_u-s\right)\right]-p_uH\left(p_l-p_u-s\right)$$

Differentiating with respect to s, imposing $s=0$ and using the above results on $\frac{\partial H(p_l-p_u-s)}{\partial s}$, $\frac{\partial p_l}{\partial s}$ and $\frac{\partial p_u}{\partial s}$, we find the derivative around $s=0$ of a locked consumer is $\frac{2}{3}$, hence the price charged to young consumers falls by $\frac{2\delta_f}{3}$. ■

4.9.3 Proofs for the No Discrimination Model

Method of Undetermined Coefficients

Maximising $\pi_A^t\left(p_A^t,p_B^t,\tilde{y}^{t-1}\right)+\delta_f V_A^{t+1}\left(\tilde{y}^t\right)$ in (4.15) with respect to p_A^t gives a first order condition

$$D_A^t\left(p_A^t,p_B^t,\tilde{y}^{t-1}\right)-p_A^th\left(\cdot\right)+\frac{\partial\tilde{y}^t}{\partial p_A^t}\left[p_A^th\left(\tilde{y}^t\right)+\delta_fN+\delta_f2R\tilde{y}^t\right]=0 \tag{4.24}$$

(I check that profit is quasiconcave later on, once K, N and R can be bounded). We can use (4.13), (4.9) and (4.10) to substitute in for $\tilde{y}$, $\frac{\partial\tilde{y}^t}{\partial p_A^t}$, p_A^t and p_B^t. Collecting terms (only addition and subtraction) gives an expression $\alpha_1+\alpha_2\tilde{y}^{t-1}=0$, and setting $\alpha_1=\alpha_2=0$

gives the condition

$$\alpha_1 = 1 - \frac{J}{D}\left[\frac{1+\delta_c K\tilde{s}}{1+2\delta_c K\tilde{s}}\right] - \frac{\delta_f N}{1+2\delta_c K\tilde{s}} = 0 \tag{4.25}$$

$$\alpha_2 = \frac{\tilde{s}}{2D} - 3\frac{K}{D}\left[\frac{1+\delta_c K\tilde{s}}{1+2\delta_c K\tilde{s}}\right] + \frac{4RK\delta_f}{\left[1+2\delta_c K\tilde{s}\right]^2} = 0 \tag{4.26}$$

To find an expression for A's period-t valuation, take $\pi_A^t\left(p_A^t, p_B^t, \tilde{y}^{t-1}\right) + \delta_f V_A^{t+1}\left(\tilde{y}^t\right)$ and replace the term $V_A^{t+1}\left(\tilde{y}^t\right)$ using (4.11). Next substitute for $\tilde{y}^t$ using (4.13) and finally use (4.9) and (4.10) to eliminate p_A^t and p_B^t from the expression. Collecting terms (only addition and subtraction), we can write A's period-t valuation as $\alpha_3 + \alpha_4\tilde{y}^{t-1} + \alpha_5\left(\tilde{y}^{t-1}\right)^2$. Since we assumed in (4.11) that this value function equals $M + N\tilde{y}^{t-1} + R\left(\tilde{y}^{t-1}\right)^2$, we can equate coefficients and get three equations

$$\alpha_3 = J + \delta_f M = M \tag{4.27}$$

$$\alpha_4 = \frac{J}{D}\frac{\tilde{s}}{2} - \frac{J}{D}2K\frac{1+\delta_c K\tilde{s}}{1+2\delta_c K\tilde{s}} + K - \frac{2K\delta_f N}{1+2\delta_c K\tilde{s}} = N \tag{4.28}$$

$$\alpha_5 = \frac{K\tilde{s}}{2D} - 2\frac{K^2}{D}\frac{1+\delta_c K\tilde{s}}{1+2\delta_c K\tilde{s}} + \frac{4K^2\delta_f R}{\left[1+2\delta_c K\tilde{s}\right]^2} = R \tag{4.29}$$

Now combine two of the equations to gain an expression for K. Notice that if $\tilde{s} > 0$ then from (4.26) $K \neq 0$. Therefore rearrange (4.26) to get

$$R = \frac{3}{4\delta_f D}\left[1+\delta_c K\tilde{s}\right]\left[1+2\delta_c K\tilde{s}\right] - \frac{\tilde{s}}{8K\delta_f D}\left[1+2\delta_c K\tilde{s}\right]^2 \tag{4.30}$$

and then substitute this into (4.29), cancel some terms, multiply by $4K\delta_f D\left[1+2\delta_c K\tilde{s}\right] \neq$

0, and then rearrange to find:

$$\frac{\tilde{s}}{2}\left[1+2\delta_c K\tilde{s}\right]^3 - 3K\left[1+\delta_c K\tilde{s}\right]\left[1+2\delta_c K\tilde{s}\right]^2 + 4\delta_f K^3\left[1+\delta_c K\tilde{s}\right] = 0 \tag{4.31}$$

If we let $\phi(K)$ be the lefthand side of (4.31), then we can write

- $\phi\left(-\frac{1}{\delta_c\tilde{s}}\right) = -\frac{\tilde{s}}{2} < 0$
- $\phi\left(-\frac{1-\tilde{s}}{2}\right) > 0$ [29]
- $\phi(0) = \frac{\tilde{s}}{2} > 0$
- $\phi\left(\frac{\tilde{s}}{6}\right) > 0$ [30]
- $\phi\left(\frac{\tilde{s}}{5}\right) < 0$ [31]
- $\phi\left(\frac{1}{2}\right) < 0$ [32]
- $\phi(\pm\infty) = \infty$ provided that δ_f is not too small compared to δ_c.[33]

[29] $\phi\left(-\frac{1-\tilde{s}}{2}\right) = \frac{\tilde{s}}{2}\left[1-\delta_c\tilde{s}(1-\tilde{s})\right]^3 + (1-\tilde{s})\frac{2-\delta_c\tilde{s}(1-\tilde{s})}{4}\left\{3\left[1-\delta_c\tilde{s}(1-\tilde{s})\right]^2 - \delta_f(1-\tilde{s})^2\right\}$

[30] $\phi\left(\frac{\tilde{s}}{6}\right) > \frac{\tilde{s}}{2}\left[1+\frac{\delta_c\tilde{s}^2}{3}\right]^3 - \frac{\tilde{s}}{2}\left[1+\frac{\delta_c\tilde{s}^2}{6}\right]\left[1+\frac{\delta_c\tilde{s}^2}{3}\right]^2 > 0$

[31] $\phi\left(\frac{\tilde{s}}{5}\right) = \frac{\tilde{s}}{2}\left[1+\frac{2\delta_c\tilde{s}^2}{5}\right]^3 - \frac{71\tilde{s}}{125}\left[1+\frac{\delta_c\tilde{s}^2}{5}\right]\left[1+\frac{\delta_c\tilde{s}^2}{5}\right]^2 - \frac{4\tilde{s}}{125}\left[1+\frac{\delta_c\tilde{s}^2}{5}\right]\left[1+\frac{2\delta_c\tilde{s}^2}{5}\right]^2 + \frac{4\delta_f\tilde{s}^3}{125}\left[1+\frac{\delta_c\tilde{s}^2}{5}\right]$.
The last two terms are negative; the first two are also negative if $1+\frac{2\delta_c\tilde{s}^2}{5} < \frac{142}{125}\left[1+\frac{\delta_c\tilde{s}^2}{5}\right]$, and this clearly holds since $\delta_c\tilde{s}^2 < \frac{85}{108}$.

[32] $\phi\left(\frac{1}{2}\right) = \frac{\tilde{s}}{2}\left[1+\delta_c\tilde{s}\right]^3 - \left[1+\frac{\delta_c\tilde{s}}{2}\right]\left[1+\delta_c\tilde{s}\right]^2 - \frac{1}{2}\left[1+\frac{\delta_c\tilde{s}}{2}\right]\left[1+\delta_c\tilde{s}\right]^2 + \frac{\delta_f}{2}\left[1+\frac{\delta_c\tilde{s}}{2}\right]$
The last two terms are negative, as are the first two because $\frac{\tilde{s}}{2}\left[1+\delta_c\tilde{s}\right] < 1+\frac{\delta_c\tilde{s}}{2}$.

[33] The quadratic term dominates when $|K|$ becomes large; it is equal to $4K^4\delta_c\tilde{s}\left[\delta_f - 3\delta_c^2\tilde{s}^2\right]$ and is therefore positive if $\delta_f > 3\delta_c^2\tilde{s}^2$; since $\tilde{s} \le \frac{1}{2}$, it holds if $\delta_f > \frac{3\delta_c^2}{4}$.

Since $\phi(K)$ is continuous in K and quartic, we can conclude that $\phi(K) = 0$ has four roots, one in each of the intervals $\left(-\infty, -\frac{1}{\delta_c \tilde{s}}\right)$, $\left(-\frac{1}{\delta_c \tilde{s}}, -\frac{1-\tilde{s}}{2}\right)$, $\left(\frac{\tilde{s}}{6}, \frac{\tilde{s}}{5}\right)$ and $\left(\frac{1}{2}, \infty\right)$.

Now solve for the other unknowns as a function of K. To find R, rewrite (4.26) and then substitute in from (4.29) to get the following:

$$\begin{aligned} R &= \frac{1}{4\delta_f D K \left[1 + 2\delta_c K \tilde{s}\right]} \left[3K \left[1 + \delta_c K \tilde{s}\right] \left[1 + 2\delta_c K \tilde{s}\right]^2 - \frac{\tilde{s}}{2} \left[1 + 2\delta_c K \tilde{s}\right]^3\right] \\ &= \frac{K^2}{D} \frac{1 + \delta_c K \tilde{s}}{1 + 2\delta_c K \tilde{s}} \end{aligned} \tag{4.32}$$

It is also simple to rearrange (4.27) and get the expression for M given in (4.19). By rearranging (4.25) we find that $J = D\left[1 + \frac{\delta_c K \tilde{s} - \delta_f N}{1 + \delta_c K \tilde{s}}\right]$, and by substituting this into (4.28) we get an expression

$$\left[1 + \frac{\delta K \tilde{s} - \delta N}{1 + \delta K \tilde{s}}\right] \left[\frac{\tilde{s}}{2} - 2K \frac{1 + \delta K \tilde{s}}{1 + 2\delta K \tilde{s}}\right] + K - \frac{2K\delta N}{1 + 2\delta K \tilde{s}} = N$$

which after simplifying can be rewritten as

$$N = \frac{\frac{\tilde{s}}{2}\left[1 + 2\delta_c K \tilde{s}\right] - K\left[1 + \delta_c K \tilde{s}\right]}{1 + \delta_c K \tilde{s} + \frac{\tilde{s}}{2}\delta_f} \tag{4.33}$$

Finally we can substitute this into $J = D\left[1 + \frac{\delta_c K \tilde{s} - \delta_f N}{1 + \delta_c K \tilde{s}}\right]$ and get the expression for J given in the text.

Properties of Equilibrium Prices

Proof of Summary 30: I only provide selected proofs. It is useful to first place bounds on the sizes of some unknown coefficients. Firstly it is clear that $RD < K^2$ and since $K < \frac{\tilde{s}}{5} \leq \frac{1}{10}$, we can conclude that $RD \in \left[0, \frac{1}{100}\right]$. Secondly N decreases in δ_f therefore $N \leq N|_{\delta_f=0} = \frac{\tilde{s}}{2}\frac{1+2\delta_c K\tilde{s}}{1+\delta_c K\tilde{s}} - K$. Since $K > \frac{\tilde{s}}{6}$ it follows that $N < \frac{\tilde{s}}{2}\frac{1+2\delta_c K\tilde{s}}{1+\delta_c K\tilde{s}} - \frac{\tilde{s}}{6}$ and since the latter is increasing in $\delta_c K\tilde{s}$ (which never exceeds $\frac{1}{20}$) we conclude $N < \frac{5}{14}\tilde{s}$. These two results are useful when proving quasiconcavity of the value function. Thirdly, $\frac{J}{D}$ increases in δ_c and falls in δ_f (see later), therefore $\frac{J}{D} > \frac{J}{D}\big|_{\delta_c=0,\delta_f=1}$; the latter is increasing in K, and once substituting for $K = \frac{\tilde{s}}{6}$, is decreasing in $\tilde{s}$ therefore $\frac{J}{D} > \frac{J}{D}\big|_{\delta_c=0,\delta_f=1,K=\frac{\tilde{s}}{6},\tilde{s}=\frac{1}{2}} = \frac{13}{15}$. Fourthly $\frac{J}{D} \leq \frac{J}{D}\big|_{\delta_f=0} = \frac{1+2\delta_c K\tilde{s}}{1+\delta_c K\tilde{s}}$ and the latter is decreasing in $\delta_c K\tilde{s}$, therefore $\frac{J}{D} < \frac{J}{D}\big|_{\delta_f=0,\delta_c K\tilde{s}=\frac{1}{20}} = \frac{22}{21}$. *It is then clear that prices are always positive.* $V_A^t = M + N\tilde{y}^{t-1} + R\left(\tilde{y}^{t-1}\right)^2$ is convex in $\tilde{y}^{t-1}$ since $R > 0$; the derivative of V_A^t with respect to $\tilde{y}^{t-1}$ is $N + 2R\tilde{y}^{t-1} \geq N - 2RD$ which is positive because

$$\begin{aligned} N &= \frac{\frac{\tilde{s}}{2}\left[1+2\delta_c K\tilde{s}\right] - K\left[1+\delta_c K\tilde{s}\right]}{1+\delta_c K\tilde{s}+\frac{\tilde{s}}{2}\delta_f} \\ &> \frac{\tilde{s}\left(\frac{1}{2}-\frac{1}{5}\right)\left(1+\delta_c K\tilde{s}\right)}{1+\delta_c K\tilde{s}+\frac{1}{4}} > \frac{6\tilde{s}}{25} > 2K^2 > 2RD \end{aligned}$$

So it follows that value increases in installed base. However even a firm with no installed base has a value $M - ND + RD^2$ and this is strictly positive since we know $M \geq J > \frac{13}{15}D$ whilst $N < \frac{5}{14}\tilde{s} \leq \frac{5}{28}$. *So a firm's value is always positive.*

Proof of Equation (4.22): Using (4.14), A's demand is:

$$\begin{aligned}
D_A^t &= H\left(\tilde{y}^{t-1}\right) H\left(p_B^t - p_A^t + s\right) + \left[1 - H\left(\tilde{y}^{t-1}\right)\right] H\left(p_B^t - p_A^t - s\right) + H\left(\tilde{y}^t\right) \\
D_A^t &= H\left(p_B^t - p_A^t - s\right) + H\left(\tilde{y}^{t-1}\right)\left[\frac{2s}{2D}\right] + H\left(\frac{p_B^t - p_A^t}{1 + 2\delta_c K \frac{s}{D}}\right) \\
D_A^t &= \frac{D - 2K\tilde{y}^{t-1} - s}{2D} + \frac{D + \tilde{y}^{t-1}}{2D}\left[\frac{s}{D}\right] + \frac{D + \frac{-2K\tilde{y}^{t-1}}{1+2\delta_c K\tilde{s}}}{2D} \\
D_A^t &= \frac{1}{2} - \frac{K\tilde{y}^{t-1}}{D} - \frac{s}{2D} + \frac{s}{2D} + \frac{s\tilde{y}^{t-1}}{2D^2} + \frac{1}{2} - \frac{\frac{K}{D}\tilde{y}^{t-1}}{1 + 2\delta_c K\tilde{s}} \\
D_A^t &= 1 - \frac{K\tilde{y}^{t-1}}{D} + \frac{s\tilde{y}^{t-1}}{2D^2} - \frac{\frac{K}{D}\tilde{y}^{t-1}}{1 + 2\delta_c K\tilde{s}} \\
D_A^t &= 1 + \frac{\tilde{y}^{t-1}}{D}\left[\frac{s}{2D} - K - \frac{K}{1 + 2\delta_c K\tilde{s}}\right] \\
D_A^t &= 1 + \frac{\tilde{y}^{t-1}}{D}\left[\frac{\tilde{s}}{2} - 2K\frac{1 + \delta_c K\tilde{s}}{1 + 2\delta_c K\tilde{s}}\right] \qquad (4.34)
\end{aligned}$$

where the third line uses (4.9) and (4.10) to rewrite $p_B^t - p_A^t$ as $-2K\tilde{y}^{t-1}$. Note that since $K \in \left[\frac{\tilde{s}}{6}, \frac{\tilde{s}}{5}\right]$, $\frac{\tilde{s}}{2} - 2K\frac{1+\delta_c K\tilde{s}}{1+2\delta_c K\tilde{s}} > 0$. Changing the time superscripts in (4.34), we find

$$\begin{aligned}
D_A^{t-1} &= 1 + \frac{\tilde{y}^{t-2}}{D}\left[\frac{\tilde{s}}{2} - 2K\frac{1 + \delta_c K\tilde{s}}{1 + 2\delta_c K\tilde{s}}\right] \\
&\Longrightarrow \quad \frac{\tilde{y}^{t-2}}{D}\left[\frac{\tilde{s}}{2} - 2K\frac{1 + \delta_c K\tilde{s}}{1 + 2\delta_c K\tilde{s}}\right] = D_A^{t-1} - 1 \qquad (4.35)
\end{aligned}$$

Using (4.9), (4.10) and (4.13) we know that $\tilde{y}^{t-1} = \frac{-2K}{1+2\delta_c K\tilde{s}}\tilde{y}^{t-2}$. Substituting this into (4.34) gives

$$D_A^t = 1 - \frac{2K}{1 + 2\delta_c K\tilde{s}}\frac{\tilde{y}^{t-2}}{D}\left[\frac{\tilde{s}}{2} - 2K\frac{1 + \delta_c K\tilde{s}}{1 + 2\delta_c K\tilde{s}}\right] \qquad (4.36)$$

and then substituting in (4.35) gives:

$$\begin{aligned} D_A^t &= 1 - \frac{2K}{1+2\delta_c K\tilde{s}}\left[D_A^{t-1} - 1\right] \\ \frac{D_A^t}{2} - \frac{1}{2} &= -\frac{2K}{1+2\delta_c K\tilde{s}}\left[\frac{D_A^{t-1}}{2} - \frac{1}{2}\right] \\ sh_A^t - \frac{1}{2} &= -\frac{2K}{1+2\delta_c K\tilde{s}}\left[sh_A^{t-1} - \frac{1}{2}\right] \end{aligned}$$

where the last line uses the fact that in each period, there is a unit mass of young consumers and a unit mass of old consumers, and therefore $sh_A^t = \frac{D_A^t}{2}$. ■

Proof of Lemma 31: The proofs are simple but messy, so I only show that prices are decreasing in δ_f. Proofs for the other two results follow a similar pattern. By inspection $\phi(K)$ increases in δ_f and therefore if $\frac{J}{D} + K$ decreases in δ_f, then every price observed in equilibrium must also decrease in δ_f. Using the expression for $\frac{J}{D}$ in (4.18), we can show that $\frac{J}{D} + K$ decreases in δ_f provided that $\frac{\partial K}{\partial \delta_f} < \frac{E}{F}$ where

$$\begin{aligned} E &= \left[1 + 2\delta_c K\tilde{s}\right]\frac{\tilde{s}}{2} - K\left[1 + \delta_c K\tilde{s}\right] \\ F &= \left[2\delta_c\tilde{s} + \delta_f\right]\left[1 + \frac{\tilde{s}}{2}\delta_f\right] - \delta_c\tilde{s} + \left[1 + \delta_c K\tilde{s} + \frac{\tilde{s}}{2}\delta_f\right]^2 \end{aligned}$$

and $E, F > 0$. Clearly F increases in K, δ_c, δ_f and $\tilde{s}$. Therefore we know that $F < F|_{K=\frac{1}{10},\tilde{s}=\frac{1}{2},\delta_f=\delta_c=1} = 3.69$. Since E is decreasing in K, we know that $E > E|_{K=\frac{\tilde{s}}{5}} = \tilde{s}\left[\frac{3}{10} + \frac{4}{25}\delta_c\tilde{s}^2\right]$. Therefore we can conclude that $\frac{E}{F} > \frac{\tilde{s}}{3.69}\left[\frac{3}{10} + \frac{4}{25}\delta_c\tilde{s}^2\right]$.

We can differentiate (4.31) with respect to δ_f and show that $\frac{\partial K}{\partial \delta_f} = \frac{G}{H}$ where

$$\begin{aligned} G &= 4K^3\left[1+\delta_c K\tilde{s}\right] \\ H &= 3\delta_c\tilde{s}\left[1+2\delta_c K\tilde{s}\right]^2\left[K-\tilde{s}\right]+3\left[1+\delta_c K\tilde{s}\right]\left[1+2\delta_c K\tilde{s}\right]\left[1+6\delta_c K\tilde{s}\right]-I \\ I &= 4\delta_f K^2\left[3+4\delta_c K\tilde{s}\right] \end{aligned}$$

We can show that $H+I$ increases in K therefore $H+I > H+I|_{K=\frac{\tilde{s}}{6}}$; $H+I|_{K=\frac{\tilde{s}}{6}}$ is increasing in $\tilde{s}$ therefore $H+I>H+I|_{K=\frac{\tilde{s}}{6},\tilde{s}=0}$=3. Then note that $I<I|_{K=\frac{1}{10},\tilde{s}=\frac{1}{2},\delta_f=\delta_c=1}=\frac{16}{125}$ therefore $H>\frac{359}{125}$. Since G increases in K, we know that $G < G|_{K=\frac{\tilde{s}}{5}} = \frac{4}{125}\tilde{s}^3\left[1+\frac{\delta_c\tilde{s}^2}{5}\right]$. Therefore we can conclude that $\frac{G}{H} < \frac{4}{359}\tilde{s}^3\left[1+\frac{\delta_c\tilde{s}^2}{5}\right]$. We want to prove that $\frac{\partial K}{\partial \delta_f} = \frac{G}{H} < \frac{E}{F}$; clearly it is therefore sufficient to show that

$$\frac{4}{359}\tilde{s}^3\left[1+\frac{\delta_c\tilde{s}^2}{5}\right] < \frac{\tilde{s}}{3.69}\left[\frac{3}{10}+\frac{4}{25}\delta_c\tilde{s}^2\right]$$

and this is easily seen to hold. ■

Proof that Switching Costs Reduce Steady State Price

Proof of Lemma 32: Begin by proving that $J < D$ whenever $\tilde{s} \leq \frac{3\delta_f}{2\delta_c}$. Take (4.18) and set $\frac{J}{D} < 1$; rearranging gives a condition $K < \frac{\frac{\tilde{s}}{2}\delta_f}{\delta_c\tilde{s}+\delta_f}$. Since $K < \frac{\tilde{s}}{5}$ it is sufficient that $\tilde{s} < \frac{3\delta_f}{2\delta_c}$.

Next show that J is monotonically falling in s when $\delta_c = \delta_f$. Impose the latter and

differentiate (4.18) with respect to s, to get that $\frac{\partial\left(\frac{J}{D}\right)}{\partial\tilde{s}} < 0$ requires $\frac{\partial K}{\partial\tilde{s}} < \frac{A}{B}$ where

$$\begin{aligned} A &= (1 + 2\delta K\tilde{s} + \delta K)\left(K + \frac{1}{2}\right) - 2K\left(1 + \delta K\tilde{s} + \frac{\tilde{s}}{2}\delta\right) \\ B &= 1 + \tilde{s} + \delta\tilde{s}^2 + \frac{\delta\tilde{s}}{2} \end{aligned}$$

and $A, B > 0$. We can show that A decreases in K, therefore $A > A|_{K=\frac{\tilde{s}}{5}}$; we can also show that $A|_{K=\frac{\tilde{s}}{5}}$ decreases in $\tilde{s}$, therefore $A > A|_{K=\frac{\tilde{s}}{5},\tilde{s}=\frac{1}{2}} = \frac{2}{5} + \delta\frac{3}{50}$. Clearly B increases in $\tilde{s}$ therefore $B < B|_{\tilde{s}=\frac{1}{2}}$. We can then conclude that $\frac{A}{B} > \frac{\frac{2}{5}+\delta\frac{3}{50}}{\frac{3}{2}+\frac{\delta}{2}}$ where the righthand side is decreasing in δ. Therefore substituting in $\delta = 1$, we conclude that $\frac{A}{B} > \frac{23}{100}$.

Then take (4.31), impose $\delta_f = \delta_c$. Differentiating with respect to $\tilde{s}$, we find $\frac{\partial K}{\partial\tilde{s}} = \frac{C}{D}$ where

$$\begin{aligned} C &= \frac{1}{2}\left[1 + 2\delta K\tilde{s}\right]^3 + 3\tilde{s}\delta K\left[1 + 2\delta K\tilde{s}\right]^2 - 3K^2\delta\left[1 + 2\delta K\tilde{s}\right]^2 \\ &\quad -12\delta K^2\left[1 + \delta K\tilde{s}\right]\left[1 + 2\delta K\tilde{s}\right] + 4\delta^2K^4 \\ D &= -3\delta\tilde{s}^2\left[1 + 2\delta K\tilde{s}\right]^2 + 3\left[1 + \delta K\tilde{s}\right]\left[1 + 2\delta K\tilde{s}\right]^2 + 3K\delta\tilde{s}\left[1 + 2\delta K\tilde{s}\right]^2 \\ &\quad +12K\delta\tilde{s}\left[1 + \delta K\tilde{s}\right]\left[1 + 2\delta K\tilde{s}\right] - 12\delta K^2\left[1 + \delta K\tilde{s}\right] - 4\delta^2K^3\tilde{s} \end{aligned}$$

and $C, D > 0$. We can show that C increases in K therefore $C < C|_{K=\frac{\tilde{s}}{5}}$; we can also show that $C|_{K=\frac{\tilde{s}}{5}}$ is increasing in $\tilde{s}$ and δ, therefore $C < C|_{K=\frac{\tilde{s}}{5},\tilde{s}=\frac{1}{2},\delta=1} = \frac{269}{400}$. We can also show that D increases in K therefore $D > D|_{K=\frac{\tilde{s}}{6}}$ (since $K > \frac{\tilde{s}}{6}$); we can also show that $D|_{K=\frac{\tilde{s}}{6}}$ is increasing in $\tilde{s}$ therefore $D > D|_{K=\frac{\tilde{s}}{6},\tilde{s}=0} = 3$. So $\frac{C}{D} < \frac{269}{1200}$.

Summing up,

$$\frac{\partial K}{\partial \tilde{s}} = \frac{C}{D} < \frac{269}{1200} < \frac{23}{100} < \frac{A}{B}$$

therefore $\frac{J}{D}$ decreases in $\tilde{s}$ when $\delta_c = \delta_f$. ■

Proof of Average and Maximum Prices

Proof of Lemma 33: If we differentiate (4.31) with respect to $\tilde{s}$ and then impose $\tilde{s} = K = 0$, we find $\frac{\partial K}{\partial \tilde{s}}\big|_{\tilde{s}=0} = \frac{1}{6}$. If we differentiate (4.18) with respect to $\tilde{s}$ we find $\frac{\partial (J/D)}{\partial \tilde{s}}\Big|_{\tilde{s}=0} = \delta_f \left[\frac{\partial K}{\partial \tilde{s}} - \frac{1}{2}\right] = -\frac{\delta_f}{3}$. So $\frac{\partial (J+DK)}{\partial \tilde{s}} = D\left[-\frac{\delta_f}{3} + \frac{1}{6}\right]$ and using (4.23), the derivative of the average price around $\tilde{s} = 0$ is just $\frac{\partial (J/D)}{\partial \tilde{s}} < 0$. ■

Proof of Lemma 34: To have $p_{A,t}, p_{B,t} < D$ it is sufficient that $J + DK < D$. Using (4.18), imposing $\delta_c = \delta_f = \delta$ and rearranging, we get a condition $\frac{2K}{\tilde{s} - 3\tilde{s}K - 2K - 2K^2\tilde{s}} < \delta$. This is harder to satisfy when K is larger, and we know $K < \frac{\tilde{s}}{5}$; substituting in for $K = \frac{\tilde{s}}{5}$ gives the desired expression. The average price (4.23) (when $\tilde{y}^{t-1} = \pm D$) is below D provided that $\frac{J}{D} + K\left[\frac{\tilde{s}}{2} - 2K\frac{1+\delta K\tilde{s}}{1+2\delta K\tilde{s}}\right] < 1$. Noting that $\delta < 1$, $K \geq \frac{\tilde{s}}{6}$ and $\tilde{s} \leq \frac{1}{2}$, we can show $\frac{\tilde{s}}{2} - 2K\frac{1+\delta K\tilde{s}}{1+2\delta K\tilde{s}} < \frac{7}{78}$. And using (4.18) and imposing $\delta_f = \delta_c = \delta$, $\frac{J}{D} + K\frac{7}{78} < 1$ is equivalent to $\frac{7K}{39\tilde{s} - 78K\tilde{s} - 78K - 7K^2\tilde{s} - \frac{7}{2}K\tilde{s}} < \delta$. The lefthand side is again increasing in K, so substituting in for $K = \frac{\tilde{s}}{5}$ gives the desired expression. ■

Comparisons of Pricing Regimes

The manipulations are lengthy but straightforward and therefore omitted. Expressions for price, profits and surplus are found in Example 24 for Partial Discrimination and in

Example 29 for Full Discrimination. Expressions for No discrimination are as follows:

Consumer Surplus can be shown to equal the same level as with $s = 0$, plus a term

$$\delta_c \frac{s^2}{4D} - \delta_c \frac{s}{2} - (J - D)(1 + \delta_c)$$

whilst per-firm profits on each cohort of consumers is $\frac{1}{2} J (1 + \delta_f)$.

Proving that Profit is Quasiconcave

To demonstrate sufficiency of A's first order condition,we need to show that $\pi_A^t(p_A^t, p_B^t, \tilde{y}^{t-1}) + \delta_f V_A^{t+1}(\tilde{y}^t)$ is quasiconcave in p_A^t around $J + K\tilde{y}^{t-1}$. A's demand comes from three sources: young, 'locked' and 'unlocked' consumers.[34] In period t, profit from selling to some (but not all) locked consumers is concave in p_A^t since $h(d)$ is uniform; the same is true about the profits made by selling to some (but not all) unlocked consumers. The value from selling to period-t young consumers is $p_A^t H(\tilde{y}^t) + \delta_f \left[M + N\tilde{y}^t + R(\tilde{y}^t)^2 \right]$; the second derivative with respect to p_A^t is then $-2h(\tilde{y}^t)[1 + 2\delta_c K\tilde{s}]^{-1} + 2\delta_f R[1 + 2\delta_c K\tilde{s}]^{-2}$. So the value from selling to some (but not all) young consumers is concave provided $2\delta_f RD < 1 + 2\delta_c K\tilde{s}$, which clearly holds since $RD < K^2$. However we also need to check quasiconcavity when price deviations are larger - when for example A sells to nobody in a specific group.

Upward price deviations When $p_A^t \in (J + K\tilde{y}^{t-1}, p_B^t + D - s)$ then A sells to all

[34] Locked consumers are old people who previously bought from A, whilst unlocked consumers previously bought from B.

three groups; when $p_A^t \in [p_B^t + D - s, p_B^t + D + 2\delta_c K s)$ then A sells only to locked and young consumers; when $p_A^t \in [p_B^t + D + 2\delta_c K s, p_B^t + D + s)$ then A sells only to locked consumers; when $p_A^t \geq p_B^t + D + s$ then A sells to nobody.

Firm value strictly decreases in p_A^t provided $p_A^t \in (J + K\tilde{y}^{t-1}, p_B^t + D - s)$. We wish to show that value is strictly decreasing in price for $p_A^t = p_B^t + D - s$. A's value is

$$H\left(\tilde{y}^{t-1}\right) p_A^t H\left(p_B^t - p_A^t + s\right) + p_A^t H\left(\tilde{y}^t\right) + \delta_f \left[M + N\tilde{y}^t + R\left(\tilde{y}^t\right)^2\right]$$

Derivative of locked-consumer profit is proportional to $H\left(p_B^t - p_A^t + s\right) - p_A^t h\left(p_B^t - p_A^t + s\right)$; if $p_A^t = p_B^t + D - s$, then this is negative provided $p_B^t + D > 3s$ which is easily shown to hold (since by assumption B follows the equilibrium strategy, therefore $p_B^t \geq J - KD$). The derivative of the value from young consumers is proportional to $H\left(\tilde{y}^t\right)\left[1 + 2\delta_c K\tilde{s}\right] - p_A^t h\left(\tilde{y}^t\right) - \delta_f N - 2\delta_f R\tilde{y}^t$. We can again impose $p_A^t = p_B^t + D - s$ and show that a sufficient condition for it to be negative is again that $p_B^t + D > 3s$. It then follows that profit is strictly decreasing in p_A^t in the whole interval $[p_B^t + D - s, p_B^t + D + 2\delta_c K s)$. It is also obvious that the same is true in the interval $[p_B^t + D + 2\delta_c K s, p_B^t + D + s)$ and A also earns zero profit when $p_A^t \geq p_B^t + D + s$. Therefore profit is quasiconcave whenever $p_A^t > J + K\tilde{y}^{t-1}$.

Downward price deviations When $p_A^t \in (p_B^t + s - D, J + K\tilde{y}^{t-1})$ then A sells to some in each group; when $p_A^t \in (p_B^t - D - 2\delta_c K s, p_B^t + s - D]$ then A sells to all its locked consumers, and to some (but not everyone) in the young and unlocked groups; when $p_A^t \in (p_B^t - s - D, p_B^t - D - 2\delta_c K s]$ then A sells to all young and locked consumers,

and to some (but not everyone) in the unlocked group; when $p_A^t \leq p_B^t - s - D$ then A sells to everyone.[35]

A's value is strictly increasing in p_A^t provided $p_A^t \in (p_B^t + s - D, J + K\tilde{y}^{t-1})$. We wish to show that value is strictly increasing in price for $p_A^t = p_B^t + s - D$. A's value is

$$p_A^t \left[H\left(\tilde{y}^{t-1}\right) + \left[1 - H\left(\tilde{y}^{t-1}\right)\right] H\left(p_B^t - p_A^t - s\right)\right] + p_A^t H\left(\tilde{y}^t\right) + \delta_f \left[M + N\tilde{y}^t + R\left(\tilde{y}^t\right)^2\right]$$

Profits earned on locked consumers are clearly increasing. The derivative of profit on unlocked consumers is proportional to $H\left(p_B^t - p_A^t - s\right) - p_A^t h\left(\cdot\right)$; by substituting in for $p_A^t = p_B^t + s - D$ we can show that this derivative is positive provided that $p_B^t < 3\left(D - s\right)$ and this clearly holds (since $p_B^t < J + KD$, $K < \frac{1}{10}$ and $s < \frac{1}{2}$). The derivative of value on young consumers is proportional to $\left[1 + 2\delta_c K\tilde{s}\right] H\left(\tilde{y}^t\right) - p_A^t h\left(\tilde{y}^t\right) - \delta_f N - 2\delta_f R\tilde{y}^t$; by substituting in for $p_A^t = p_B^t + s - D$ and using the facts that $R < K^2 < \frac{1}{100}$, $N < \frac{5}{14}$, $\tilde{s} < \frac{1}{2}$, $\delta_c \geq 0$ and $\delta_f < 1$, we can show that this is positive. Therefore value is increasing in price in the region $p_A^t \in (p_B^t - D - 2\delta_c Ks, p_B^t + s - D]$. It is also clear that profit is increasing in price in the region $p_A^t \in (p_B^t - s - D, p_B^t - D - 2\delta_c Ks]$ - value is clearly strictly increasing on the young and locked consumers, and it is simple to prove that the same holds for unlocked consumers. When $p_A^t \leq p_B^t - s - D$ then A sells to everybody, so value is obviously also strictly increasing in p_A^t. Therefore for all prices below $J + K\tilde{y}^{t-1}$, A's value is increasing in price. Therefore profit is quasiconcave. ■

[35] Notice that some of these deviations involve negative prices and therefore negative period-t profits, and potentially negative period-t valuations as well.

Chapter 5

Price Collusion when Firms have Different Costs

Abstract: Two firms with different marginal costs bargain over a collusive price. The low-cost firm is shown to use its stronger position in the competitive market to bargain down the price. Greater patience may *weaken* a firm's bargaining power because it affects the range of incentive compatible prices. Comparative statics in patience can therefore be non-monotonic. Low-cost entry into a duopolistic market may *raise* the collusive price, because the entrant weakens the original low-cost firm's threat point. Low-cost entry into a duopolistic market can also sometimes make collusion *easier* to sustain.

5.1 Introduction

> "People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices." *Adam Smith, The Wealth of Nations*

Economists have long noted that firms have incentives to abandon competition and reach price agreements with their rivals. Collusion appears to have been widespread in a number of industries during the last century. Levenstein and Suslow [50] mention *inter alia* agriculture, chemicals, glass, machinery, paper, plastics, steel and vitamins. Modern anti-trust authorities have widespread powers to protect consumer interests by investigating and punishing price-fixing. Under European law (Article 81) for example firms can be raided and, if found guilty, fined up to ten percent of global sales. The UK Enterprise Act (2002) permits unlimited fines and five-year jail sentences for employees found to have been complicit in collusive agreements. Despite the penalties, allegations of collusion remain relatively common. In recent years there have been allegations of price-fixing concerning Scottish dairies; drug companies; toy retailers; brewers in the Dutch beer market; elevator manufacturers; replica football shirts in the UK. Several Samsung executives admitted their role in a cartel which artificially raised the price of dynamic random access memory chips over four years. The chips are key components in computers manufactured by amongst others Dell and Hewlett-Packard. Two executives from British Airways resigned from the company admitting 'inappropriate conversations' amid allegations that the company had discussed the level of its fuel surcharges with rival

airlines.

Collusion therefore appears to remain a serious problem for anti-trust organisations. The variety of industries affected and the sums of money paid out in lawsuits attests to the size of the impact price-fixing has on consumers. Friedman's [28] article on repeated games shows how collusion can be explained within a dynamic setting where firms place enough weight on future profits. However typically many possible equilibria exist, providing in Tirole's [71] words an 'embarrassment of riches'. When firms have different costs, there is no 'focal' equilibrium. Furthermore, if the firms were to try to collude *implicitly*, there is no way they could coordinate on a specific price. We therefore model a situation in which firms meet up and discuss *explicitly* a collusive price.

Our model has two firms with different marginal costs, competing in a homogeneous goods industry. Market demand is assumed to be split equally amongst firms that are tied for the lowest price. There is a connected set of incentive compatible prices, with firms using alternating offer bargaining to select one. Firms Bertrand compete in the market until agreement on a collusive price is reached. Interesting features of this bargaining game are that (1) firms have partially coincident preferences over the collusive price, and (2) the bargaining outcome must satisfy an incentive compatibility constraint in the subsequent pricing supergame.

A unique bargaining equilibrium is shown to exist, in which the efficient firm is often able to use its stronger competitive position to bargain down the collusive price. Unlike in standard bargaining games, firms may make offers which are strictly acceptable to the

other party, and therefore the usual Rubinstein indifference equations often fail to hold. The most notable comparative statics result is that the bargained collusive price may be non-monotonic in a firm's patience level. Typically one would expect that if the low-cost firm becomes more patient, its bargaining power increases and so the collusive price falls towards its monopoly price. However in our model the opposite can occur. The high-cost firm may be prevented from offering some high prices because the low-cost firm would deviate from them. Greater patience on the latter's part may render some high prices colludable, allowing the high-cost firm to offer them. We then extend the alternating offer model by allowing the time between bargaining offers and counteroffers to become small (but still considering just two firms). The model becomes more tractable, and is shown to solve a constrained Nash bargaining problem. We demonstrate for example that the collusive price is non-monotonic in the high-cost firm's marginal cost. Harrington [36] also uses Nash bargaining and shows a similar result.[1]

The most interesting applied industrial organisation contribution comes in Section 5.7 where we study entry into collusive markets. We begin by showing that entry of a relatively low-cost firm may make collusion *easier* to sustain. Although the entrant steals market share from incumbents (thus making deviation from a collusive agreement more attractive), it also reduces the efficient incumbent's Bertrand profits and therefore makes the market more symmetric.[2] We then suppose that the entrant and the two

[1] His model is more general because firms bargain over market share and price. However he does not derive it from an alternating offer model, and does not provide comparative statics in discount factors.

[2] Vasconcelos [74] studies collusion between quantity-setting firms, and finds that if a merger makes asset holdings more unequal (so costs are less symmetric) then collusion is more difficult. In our price-competition model, entry is somewhat like a reverse-merger.

incumbents use Nash Bargaining to arrive at a collusive price. We show that, surprisingly, low-cost entry can *raise* the collusive price because the entrant undermines the low-cost incumbent's competitive profit, which it used to leverage its way to a lower collusive price. Similarly high-cost entry can *reduce* the collusive price.

5.2 Related Literature

In an early analysis of the problem, Patinkin [59] likened a cartel and its members to a multi-plant monopolist. He argued that price would be set to maximise joint profits, and quotas assigned to firms in order to minimise the total cost of a given level of production. This would typically result in high cost firms not producing anything, and in the long term shutting down. Bain [9] noted that high cost firms would only adhere to this if they received a share of cartel profits in the form of side payments. Nevertheless it seems doubtful that cartel members would make payments to firms that had closed down. In practice, Bain argued, earnings would follow output. Each firm would then attempt to bargain its way to a large quota, at a price most favourable to it given its cost structure. Joint profit maximisation could only be an accidental by-product of a cartel, rather than its defining feature (as Patinkin had envisaged).

Schmalensee [63] notes Bain's criticism of assuming joint profit maximisation, and analyses a repeated Cournot game. He considers different 'collusive technologies' (e.g. quotas), and then uses axiomatic concepts such as Nash and Kalai-Smorodinsky to pick a point on the frontier corresponding to each technology. However numerical methods

are required to solve his model, and because he does not consider incentive compatibility, he implicitly assumes that firms are perfectly patient.

Friedman [28] demonstrates that in a repeated oligopoly model, there will typically exist many possible non-cooperative equilibria on the Pareto frontier, that can be sustained as a subgame perfect equilibrium. He also proposes using a 'balanced temptation equilibrium' to select among them. In such an equilibrium, the critical discount factor at which a player is just indifferent between deviating and not, is the same for all firms.[3] Bae [6] uses a selection criterion which imposes balanced temptation, and (for given market quotas) picks price to maximise joint profits. He finds that technological progress by the high cost firm can damage social welfare (even when the progress is costless). However Harrington [36] criticises Bae's approach as being too ad hoc. In a Cournot duopoly, Verboven [75] notes that if the discount factor is sufficiently low, points on the profit possibilities frontier will not be subgame perfect, so balanced temptation equilibrium does not exist. However picking quantities where both firms' incentive compatibility constraints just bind picks out a (and possibly the only) Pareto optimal point.[4]

Harrington [36] argues in favour of what he terms a more 'primitive' solution concept. He uses Nash bargaining in a Bertrand duopoly to pick a market share for each firm and a collusive price. Comparative statics in cost parameters are shown to be qualitatively different depending upon whether or not one takes account of incentive compatibility.

[3] Deviating from collusion brings a short term gain, but then a loss in each future period due to reversion to one-shot Nash play. The ratio between the short term gain and long term per period loss is then the same for all firms.

[4] Pareto optimal in the usual sense, except that the allocation must be subgame perfect. Since he considers low discount factors, points on the profit possibilities frontier are not subgame perfect.

This is therefore an important methodological point. Harrington also uses his duopoly model to show that entry by a low cost firm reduces the collusive price, and entry by a high cost firm raises it.

5.3 The Model

There are two firms, l and h, producing homogenous goods. The total cost to firm i of producing output q_i is $C_i(q_i) = c_i q_i$. We assume $c_l < c_h$ and therefore asymmetry in cost structure. We also assume that the two firms have potentially different discount rates, denoting these by $r_l, r_h \in (0, \infty)$. For notational convenience, we work primarily in terms of discount factors, where $\delta_i \equiv \exp(-r_i) \in (0, 1)$.[5] Market demand, $D(p)$, is assumed to be continuous and strictly decreasing. Its derivative $D'(p)$ also exists and is continuous.[6] For tractability, we also assume

- (A1) $\Pi_i(p) = (p - c_i)D(p)$ is concave in $p, i = l, h$.[7]

We also assume that there is a unique monopoly price for each firm, denoted by $p_i^m = \arg\max(p - c_i)D(p)$, where $p_l^m < p_h^m$. We also assume that individual firm demand satisfies the following

[5] Harrington [35] provides reasons why δ_i may vary across firms.

[6] Although we use it to simplify later proofs of comparative statics, this assumption only becomes essential when we look at limit of frictionless bargaining.

[7] This guarantees that the Pareto frontier - the set of prices between the two firms' monopoly prices - is concave. It plays a key role in our later uniqueness proofs. Concavity of the frontier is a standard assumption in most bargaining problems.

- (A2) $D_i(p_i, p_j) = \begin{cases} D(p_i) & \text{if } p_i < p_j \\ \frac{1}{2}D(p_i) & \text{if } p_i = p_j \\ 0 & \text{if } p_i > p_j \end{cases}$

The firm with the lowest price supplies all market demand, and if prices are tied demand is split equally. It is well known that the *one-shot* game has a unique Bertrand equilibrium in which h prices at its marginal cost c_h, and l undercuts this by a vanishingly small amount. l and h then earn $\Pi_l(c_h)$ and 0 respectively.

5.4 The Pricing Supergame

If the stage game is played a countably infinite number of times, Friedman [28] shows that non-Bertrand prices can be supported as part of a subgame perfect equilibrium (SPE) when firms are sufficiently patient. A strategy S_i specifies a price for firm i at each time t in the supergame, for every possible history $J(t)$. We consider *grim strategies* in which each firm prices at $\tilde{p}$ in initial period k, and also in each subsequent period provided nobody has played anything other than $\tilde{p}$ in the past. If anybody does deviate and play something other than $\tilde{p}$, the firms play one-shot Bertrand in all subsequent periods.[8]

[8] We therefore restrict attention to stationary strategies where (absent deviations) firms play the same price $\tilde{p}$ in all periods.

Proposition 38 *Grim strategies S_l, S_h induce a SPE iff*

$$\frac{1}{2}\Pi_l(\tilde{p}) \geq (1-\delta_l)\left[\max_{p_l}(p_l - c_l)D_l(p_l,\tilde{p})\right] + \delta_l\Pi_l(c_H) \quad (5.1)$$

$$\frac{1}{2}\Pi_h(\tilde{p}) \geq (1-\delta_h)\left[\max_{p_h}(p_h - c_h)D_h(\tilde{p},p_h)\right] \quad (5.2)$$

Proof: Fix S_h, and note that $\forall t\ \forall J(t)$, subgames can be split according to whether there has or has not been a deviation. In subgames where there has been a deviation, S_l clearly induces a Nash equilibrium. In subgames where there has not been a deviation, following S_l gives payoff $\sum_{s=k}^{\infty}\delta_l^{s-k}\frac{1}{2}\Pi_l(\tilde{p}) = \frac{\Pi_l(\tilde{p})}{2(1-\delta_l)}$, whilst the best one-time deviation gives $\max_{p_l}(p_l - c_l)D_l(p_l,\tilde{p})$ followed by $\sum_{s=k+1}^{\infty}\delta^{s-k}\Pi_l(c_H) = \frac{\delta_l}{1-\delta_l}\Pi_l(c_H)$. Hence S_l induces a Nash equilibrium in these subgames if (5.1) holds. (5.2) is similarly obtained for S_h. ■

When a firm deviates from S_i, it makes a short term gain, but then suffers a long term loss since it receives Bertrand profits instead of collusion at $\tilde{p}$. If the firm is sufficiently patient, it puts enough weight on the long term losses and so finds it optimal not to deviate. If c_h is large relative to c_l, l receives large Bertrand profits and therefore the long term losses following deviation are less costly. This gives l greater incentives to deviate from any $\tilde{p}$. To make the problem interesting, we assume

- (A3) There is at least one price satisfying (5.1) and (5.2)

Hence we assume that firms are sufficiently patient, and asymmetries are sufficiently small. It then follows that:

Corollary 39 *There exists a nonempty set $\Omega = [p_l^m, \hat{p}]$ of prices that satisfy (5.1) and (5.2), where $\hat{p}$ is the highest price below p_h^m which satisfies (5.1) and (5.2)*

Proof: Firstly (5.2) cannot hold if $\delta_h < \frac{1}{2}$, so by (A3) $\delta_h \geq \frac{1}{2}$. Therefore any $\tilde{p} \in \Omega$ satisfies (5.2). Secondly if $\tilde{p} = p_l^m$ does not satisfy (5.1) then no price does. So by (A3) p_l^m does satisfy (5.1). When $\tilde{p} \geq p_l^m$, (5.1) reduces to $\frac{1}{2}\Pi_l(\tilde{p}) \geq (1-\delta_l)\Pi_l(p_l^m) + \delta_l \Pi_l(c_H)$ and this is less likely to hold as $\tilde{p}$ increases. So if p_h^m solves (5.1), so does any $\tilde{p} \in [p_l^m, p_h^m]$. And if $\hat{p} < p_h^m$, no price above $\hat{p}$ satisfies (5.1). ■

There is a connected set of incentive compatible (IC) prices Ω that lies between the firms' monopoly prices. Ω becomes the bargaining set of our game. The firms never bargain to a price below p_l^m since p_l^m is always IC and is better for both parties. Similarly the firms never bargain to a price above $\hat{p}$: *either* $\hat{p} = p_h^m$ and $\hat{p}$ is preferred to all higher prices, *or* $\hat{p} < p_h^m$ and no higher prices are IC. Hence the collusive price must lie in $\Omega = [p_l^m, \hat{p}]$, with l preferring lower prices and h higher prices.

All prices in Ω are weakly below p_h^m, so h is equally tempted to deviate at all of them. Provided $\delta_h > 1/2$, small changes in h's patience level will not affect its willingness to collude at a price in Ω. However prices in Ω are weakly above p_l^m, so l's temptation to deviate is higher, the greater the collusive price. As noted above, deviating brings a short-term gain but a long-term loss. If l becomes more patient, it places more weight on the long-term, and therefore some high prices (that were not previously incentive compatible) may become colludable.

5.5 Alternating Offers Bargaining Game

The previous section characterises the set of prices Ω that are subgame perfect in the collusive supergame, and which lie between the two firms' monopoly prices. However within Ω, l prefers lower prices and h higher prices. We now model a situation in which the two firms meet and negotiate a non-binding agreement on a collusive price. We will assume that these discussions are informative as to the way the parties will actually play in the supergame.[9]

Bargaining immediately precedes the pricing game, and has the following structure. At $t = 1$ firm i makes firm j a price offer $p_i(1)$ at which they will collude for all future periods. If j accepts, they enter the pricing game (with $p_i(1)$ replacing $\tilde{p}$). If j rejects, the firms enter a *disagreement game*, in which they non-cooperatively choose a market price for that period. At $t = 2$, j then counteroffers $p_j(2)$. If i accepts, they enter the pricing game (with $p_j(2)$ replacing $\tilde{p}$), otherwise they enter another disagreement game and then i counteroffers. This repeats itself until agreement (if any) is reached. No offers can be retracted, and there is no renegotiation.

At any time t of the bargaining game, a history $H(t)$ is a sequence of past price offers and prices posted by l and h in the ensuing disagreement games. A strategy σ_i specifies for every history, either a price offer (if it is i's turn to make an offer), or an acceptance/rejection rule (if $j \neq i$ makes an offer), and a price in any ensuing disagreement game. To simplify the task of finding SPE strategies of the bargaining

[9] Clearly legal agreements are impossible, hence the need for SPE.

game, we initially assume:

- (A4) Strategies are stationary: in equilibrium firm i always makes the same offer and uses the same acceptance rule
- (A5) No delay: in equilibrium, offers are accepted
- (A6) Firms play Bertrand Nash during disagreement games

It is standard in bargaining problems to initially assume (A4) and (A5), and then relax them later on. We will follow this approach as well. (A6) is natural and will be discussed later on.

5.5.1 Best responses

This section studies what a player's optimal offer looks like.

Optimal offers must lie in Ω. By (A5) offers are acceptable, so they must be IC in the supergame. Moreover prices outside Ω are either not IC and/or are worse for both firms than some $\tilde{p} \in \Omega$. Therefore any profit maximising offer must lie in Ω.

Now take a point in the bargaining game where h makes an offer. Suppose that if l rejects, next period the firms start colluding at price $\bar{p}_l$. Then h's optimal price offer (its 'best response') is to offer $B_h(\bar{p}_l)$ where

$$B_h(\bar{p}_l) = \max_{p_h \in \Omega} \Pi_h(p_h) \text{ s.t. } \Pi_l(p_h) \geq \delta_l \Pi_l(\bar{p}_l) + 2(1-\delta_l)\Pi_l(c_h) \tag{5.3}$$

Proof: By (A5) p_h must be acceptable to l and therefore IC. Accepting $p_h \in \Omega$ gives l a payoff $\frac{1}{2(1-\delta_l)}\Pi_l(p_h)$; rejecting gives $\Pi_l(c_h) + \frac{\delta_l}{2(1-\delta_l)}\Pi_l(\bar{p}_l)$, hence the inequality in (5.3). Subject to this h chooses p_h to maximise his collusive payoff $\frac{1}{2(1-\delta_h)}\Pi_h(p_h)$. ■

Suppose we divide the inequality in (5.3) by half. The left-hand side is then l's average payoff from accepting p_h. The right-hand side is l's average payoff from rejecting p_h - namely a weighted average of Bertrand profit and collusive profit at $\bar{p}_l$. l accepts p_h if and only if accepting p_h gives it a higher average payoff than rejecting. h then offers a price that maximises its average collusive payoff, $\Pi_h(p_h)$, subject to acceptance. If $\bar{p}_l \in \Omega$ (as it will be in the later equilibrium analysis) and $B_h(\bar{p}_l) < \hat{p}$, then h's offer makes l just indifferent about accepting or rejecting - more formally

$$B_h(\bar{p}_l) = \Pi_l^{-1}(\delta_l \Pi_l(\bar{p}_l) + 2(1-\delta_l)\Pi_l(c_h)) \tag{5.4}$$

In a typical Rubinstein problem with monotonic payoff functions, a player's optimal offer will make the other just indifferent between accepting and rejecting. This is true in our model when $B_h(\bar{p}_l) < \hat{p}$, but not when $B_h(\bar{p}_l) = \hat{p}$. l may strictly prefer to accept $\hat{p}$ (rather than delay and then collude at $\bar{p}_l$) - its acceptance inequality in (5.3) is strict. In an ordinary Rubinstein problem h would raise its offer to exploit this. However in our model h does not do this, because either (a) $\hat{p} = p_h^m$ and h does not want to raise its offer or (b) $\hat{p} < p_h^m$ and h would like to raise its offer, but incentive compatibility prevents it from doing so. We can also write l's best response as (5.5), and if $\bar{p}_h \in \Omega$ and $B_l(\bar{p}_h) > p_l^m$ it can be written as (5.6).

$$\max_{p_l \in P} \Pi_l(p_l) \text{ s.t. } \Pi_h(p_l) \geq \delta_h \Pi_h(\bar{p}_h) \tag{5.5}$$

$$B_l(\bar{p}_h) = \Pi_h^{-1}\left(\delta_h \Pi_h(\bar{p}_h)\right) \tag{5.6}$$

5.5.2 Subgame Perfect Bargaining Equilibrium

Proposition 40 *A bargaining SPE satisfying our assumptions has*

- l offers $p_l^* = B_l(p_h^*)$, h offers $p_h^* = B_h(p_l^*)$
- l accepts any IC p_h s.t. $\Pi_l(p_h) \geq \delta_l \Pi_l(p_l^*) + 2(1-\delta_l)\Pi_l(c_h)$; h accepts any IC p_l such that $\Pi_h(p_l) \geq \delta_h \Pi_h(p_h^*)$
- In every disagreement game, l and h play one-shot Bertrand Nash

Proof: The strategies induce a Nash equilibrium in disagreement games, and the acceptance rules were previously shown to be optimal. Offers are optimal *provided firms find it optimal to make acceptable offers.* And they do, because delay is costly for both parties. ■

Proposition 41 *The above bargaining equilibrium always exists, and is unique*

This is due to our assumption that profit functions (and therefore also the Pareto frontier) are concave. As in standard bargaining problems, without this assumption

there may be multiple equilibria. In the present context, it ensures that best response functions are well-behaved. We also note

Corollary 42 *In the bargaining SPE,* $p_l^* < p_h^*$ *unless* $\Omega = p_l^m$

Proof is in the appendix. ■

This is very intuitive and results from a first-mover advantage. If $\Omega = p_l^m$, then the only (non-Bertrand) price that is incentive compatible in the supergame is p_l^m, so both l and h will offer this. However when Ω is non-degenerate, l prefers lower prices and h higher prices. Delay (and the associated Bertrand profits) are costly for both firms. Therefore if l can get p_l^* accepted next period, it would strictly prefer to accept an offer of p_l^* now from h. This allows h to offer a slightly higher price in Ω and l will still accept. This first-mover advantage is familiar from Rubinstein's original model.

5.5.3 Explicit solution

We have now shown that the bargaining game satisfying our assumptions has a unique equilibrium, characterised by best response functions. Best responses may be at p_l^m or $\hat{p}$ - in other words at the corners of the bargaining set Ω. This section explicitly solves the bargaining game on a case-by-case basis.

Best response functions are defined on $\Omega = [p_l^m, \hat{p}]$, therefore (a). either $B_l(\hat{p}) = p_l^m$ or $B_l(\hat{p}) > p_l^m$, and also (b). either $B_h(p_l^m) = \hat{p}$ or $B_h(p_l^m) < \hat{p}$. This gives four possible combinations, one of which will always hold for any problem. We summarise these as follows:

L1 $B_l(\hat{p}) = p_l^m$ L2 $B_l(\hat{p}) > p_l^m$

H1 $B_h(p_l^m) = \hat{p}$ H2 $B_h(p_l^m) < \hat{p}$

L1 vs. L2 Suppose that l makes h an offer, and if h rejects it, the firms collude at $\hat{p}$ (h's best price) next period. Can l get p_l^m accepted? (Does L1 hold?) If h rejects, it suffers a short-term loss (Bertrand profit of 0 versus collusion at p_l^m for one period) but has a long-term gain (collusion at $\hat{p}$ instead of p_l^m in all subsequent periods). L1 is more likely to hold if h is impatient, and $\hat{p}$ is close to p_l^m (such that the long-term gain in collusive price is small).

H1 vs. H2 Suppose now that h makes l an offer, and that if l rejects the firms collude at p_l^m (l's best price) next period. Can h get $\hat{p}$ accepted? (Does H1 hold?) Again if l rejects the offer, it suffers a short-term loss but a long-term gain. Therefore H1 is more likely to hold if l is impatient and $\hat{p}$ is close to p_l^m.

We can now start characterising the solution on a case-by-case basis.

Case 1: L1 and H1 hold[10]

The solution is $p_l^* = p_l^m$ and $p_h^* = \hat{p}$.

In Case 1, the proposer offers collusion at his preferred price, and the rival firm accepts. Our above discussion illustrates that Case 1 occurs when l and h are impatient relative to the bargaining set Ω. Firms may be impatient, placing little weight on future profits and being more concerned with reaching quick agreement (even on a relatively

[10] **For most of the Cases, proofs follow directly from Propositions 40 and 41, and so are omitted.**

unfavourable price). Alternatively Ω may be small (either because $c_h \approx c_l$ so $p_h^m \approx p_l^m$, or c_h is large so only prices near p_l^m are IC in the pricing supergame). Then it is not worth delaying, since a firm will only get collusion at a slightly more favourable price.

Case 2: L1 and H2 hold

The solution is $p_l^* = p_l^m$ and $p_h^* = \Pi_l^{-1}(\delta_l \Pi_l(p_l^m) + 2(1-\delta_l)\Pi_l(c_h))$

Proof: $B_l(\cdot)$ is non-decreasing, so L1 implies $B_l(B_h(p_l^m)) = p_l^m$. Then solve for p_h^* using (5.4) and apply Proposition 41. ■

Case 2 occurs when l is patient relative to the bargaining set, but h is not. Regardless of what h can get in the future, it is impatient and therefore prefers to accept p_l^m now rather than suffer one period's delay. However l is patient, and this forces h to offer a collusive price below $\hat{p}$ to gain acceptance. Notice that p_h^* is independent of δ_h, because it is the highest price in Ω that l will accept - h's preferences have no direct impact upon the offer it makes.

Case 3: L2 and H1 hold

L2 and H1 hold. The solution is $p_l^* = \Pi_h^{-1}(\delta_h \Pi_h(\hat{p}))$ and $p_h^* = \hat{p}$

The intuition is the same as for Case 2, except that now h is patient relative to the bargaining set and l is not. This means that l will not reject high price offers by h (which h therefore makes), but h will reject low price offers by l, forcing l to make an offer above p_l^m. Therefore p_l^* is the lowest price that h is prepared to accept, given that if h rejects it, h gets zero Bertrand profit for one period followed by collusion at its preferred price

$\hat{p}$.

The final scenario we need to consider is when L2 and H2 both hold. It turns out that these conditions by themselves are not sufficient to characterise the solution, so we introduce four new conditions:

*1 $B_h(B_l(\hat{p})) = \hat{p}$ *2 $B_h(B_l(\hat{p})) < \hat{p}$

*3 $B_l(B_h(p_l^m)) = p_l^m$ *4 $B_l(B_h(p_l^m)) > p_l^m$

Again (a) either *1 or *2 hold, and (b) either *3 or *4 hold. This appears to give four possible cases, though we show in the appendix that actually there are only three. In Cases 1-3 the form of the solution depended upon how patient a player was relative to the bargaining set Ω. If both were impatient, firms offered their preferred collusive price. If only one player was impatient relative to Ω, they were forced to make an offer that was more favourable to their rival firm in order to gain acceptance. When (L2) and (H2) hold, both firms are patient relative to Ω. The stress now is on how patient one firm is relative to their rival.

To understand the conditions, suppose that h is patient and l impatient. Suppose that l makes h an offer, and if h rejects it can get collusion at $\hat{p}$ next period. This faces h with the usual trade-off, and if h is patient, we have already noted it has stronger incentives to reject any offer by l. Therefore $B_l(\hat{p})$ - l's lowest acceptable offer - will have to be higher. Now suppose that h is making l an offer, and if l rejects the firms will collude at price $B_l(\hat{p})$ next period. l also has the familiar trade-off, and if l is impatient it will not reject high prices, meaning that $B_h(B_l(\hat{p}))$ - h's highest acceptable price - is likely to be higher. In particular, it is more likely that *1 holds and $B_h(B_l(\hat{p})) = \hat{p}$.

Similarly *2 is more likely to hold when l is relatively patient. One can use similar reasoning to show that *3 is more likely (and *4 less likely) when l is relatively more patient than h.

Case 4: *1 holds

The solution is $p_l^* = \Pi_h^{-1}(\delta_h \Pi_h(\hat{p}))$ and $p_h^* = \hat{p}$

We know that *1 is more likely to hold when h is patient relative to l. As usual, rejecting an offer brings a short-term loss but a long-term gain. If h is patient, it is willing to reject low offers, and this forces l to make a relatively high offer. l's relative impatience means that it, meanwhile, is not so willing to reject higher price offers. Together these facts imply that if h is sufficiently patient relative to l, h is able to tip the solution to $\hat{p}$. l's offer is then made to make h just willing to accept, and so as in Case 3, p_l^* is independent of δ_l.

Case 5: *3 holds

The solution is $p_l^* = p_l^m$ and $p_h^* = \Pi_l^{-1}(\delta_l \Pi_l(p_l^m) + 2(1-\delta_l)\Pi_l(c_h))$

The explanation is similar to that behind Case 4, except that now l is relatively patient as compared with h, allowing l to tip the solution to the opposite corner at p_l^m. h is not patient enough to credibly reject p_l^m given its own offer p_h^*. p_h^* in turn is set to make l just willing to accept given that if l rejects, it can get collusion at p_l^m accepted next period.

Case 6: *2 and *4 hold

$p_l^m < p_l^* < p_h^* < \hat{p}$, and p_l^*, p_h^* solve $\Pi_l(p_h^*) = \delta_l \Pi_l(p_l^*) + 2(1-\delta_l)\Pi_l(c_h)$ and $\Pi_h(p_l^*) = \delta_h \Pi_h(p_h^*)$

Proof: $p_l^* = B_l(B_h(p_l^*))$. *4 implies $p_l^* \neq p_l^m$ and *2 implies $p_h^* \neq \hat{p}$. So by Proposition 41 there is a unique interior equilibrium and it satisfies (5.4) and (5.6). ■

This is the only case in which both price offers are in the interior of the bargaining set Ω. Therefore both firms make an offer that makes the rival just indifferent between accepting and rejecting (if not, l for example could lower the price offer slightly, still get it accepted, and be better off). As a result this is the only case in which the Rubinstein indifference equations hold. This case is likely to occur when the patience levels of both firms are not too unbalanced. In particular, unlike in Cases 4 and 5, neither firm is sufficiently patient that they can use this to tip the solution to their preferred corner of the bargaining set.

Summary of Cases

The following Table gives all the solutions. Although there are six cases of interest, there are only four solution types, as one might expect. In Case 1 both price offers are at corners. In Case 6 neither price offer is at a corner. In Cases 2 and 5 l's offer is at a corner, and h's offer is interior. The reverse is true in Cases 3 and 4.

Case	$\mathbf{p}_l^*$	$\mathbf{p}_h^*$
1	p_l^m	$\hat{p}$
2	p_l^m	$\Pi_l^{-1}(\delta_l \Pi_l(p_l^m) + 2(1-\delta_l)\Pi_l(c_h))$
3	$\Pi_h^{-1}(\delta_h \Pi_h(\hat{p}))$	$\hat{p}$
4	$\Pi_h^{-1}(\delta_h \Pi_h(\hat{p}))$	$\hat{p}$
5	p_l^m	$\Pi_l^{-1}(\delta_l \Pi_l(p_l^m) + 2(1-\delta_l)\Pi_l(c_h))$
6	Solve $\Pi_l(p_h^*) = \delta_l \Pi_l(p_l^*) + 2(1-\delta_l)\Pi_l(c_h)$ and $\Pi_h(p_l^*) = \delta_h \Pi_h(p_h^*)$	

We saw that Cases 1-3 are determined by how patient firms are relative to the bargaining set Ω. In Cases 4-6 the focus is instead on how patient firms are relative to each other. If one firm is sufficiently patient relative to the other, then they tip the solution towards the corner of the bargaining set most favourable to them. In all cases there is a strong tendency for at least part of the solution to be at a corner of the bargaining set. This is due to two factors:

1. Profit functions are non-monotonic. For example if l can get p_l^m accepted, it *does not want to* make h any other offer.

2. Incentive-compatibility. In general we might have $p_h^* = \hat{p} < p_h^m$ and l strictly preferring to accept. h would like to make a higher offer, but *cannot*, since it would not be incentive-compatible in the pricing game.

Firms therefore tend to make price offers at corners either because they have no choice (price offers above $\hat{p} < p_h^m$ are not incentive compatible), or because they do not want to

due to the nature of profit functions. As a result our solution bears a strong resemblance to a game in which profit functions are *monotonic*, but there are outside options at p_l^m and $\hat{p}$. However the interpretation is completely different. With outside options, l would offer p_l^m because h either could not, or would not, accept lower prices. In our model, by contrast, h typically *would* accept slightly lower prices (recall that h usually strictly prefers to accept), but l itself chooses not to make these offers. Moreover $\hat{p}$ is influenced by δ_l in a way that outside options in a usual bargaining game would not be.[11]

5.5.4 A comment on relative bargaining strengths

The low-cost firm has an advantage in bargaining, but the high-cost firm may still have significant bargaining power.

The low-cost firm is stronger because the bargaining set may not include high prices such as p_h^m (l has too strong incentives to deviate from such prices). For example in Cases 1,3 and 4 $p_h^* = \hat{p}$ and so if $\hat{p} < p_h^m$, h is prevented from offering p_h^m. This in turn allows l to make lower (acceptable) offers. Suppose instead that the anti-trust authority was *pro-collusion* in the sense that it policed collusive agreements. If it punished deviators sufficiently hard, the authority could render prices in $(\hat{p}, p_h^m]$ incentive compatible. This would assist inefficient firms in bargaining for a higher price, and hence softer anti-trust policy could *harm* efficient firms.

[11] The solution technique for a model with outside options is different too. In such a model one would typically begin by solving the problem without outside options, using indifference equations. One would then impose the outside options.

The low-cost firm is also stronger in bargaining because it receives Bertrand profit during disagreement. This makes it less costly for l to reject an offer, and so since h always makes acceptable offers, this forces h to make an offer more favourable to l. In turn this allows l to make lower offers and still have h accept. Therefore l can leverage its superior competitive profit into the collusive market, and use this to bargain down the agreed price.

Despite this, the high-cost firm still has a certain amount of bargaining power. In a case study on Danish cartels, Fog [27] writes "Little regard is paid to the inefficient firms, simply because they have no bargaining power." For the reasons mentioned above, our model certainly shows that inefficient firms will have *less* bargaining power ceteris paribus than more efficient rivals. However unless $\Omega = p_l^m$ (the bargaining set is degenerate), h does have some power. In particular, when h makes an offer, it can achieve collusion above l's monopoly price. Indeed if Ω is large, l itself may even be forced to offer collusion above its own monopoly price (see Cases 3,4 and 6). h's power derives from the assumption that bargaining is 'fair' in the sense that both firms have to wait the same amount of time before counteroffering.[12]

[12] This is unrelated to the first-mover advantage. The next section extends the model to 'quick' counteroffers, which eliminates the first-mover advantage. However collusion still occurs at a price above p_l^m. If instead l could counteroffer much more quickly than h, l would hold all the bargaining power (as in standard bargaining models), because rejecting an offer would be almost costless for it (compared to h's cost of rejecting an offer).

5.5.5 Comparative Statics

δ_h

Proposition 43 *p_l^*, p_h^* are (weakly) increasing in δ_h*

In a standard bargaining model, an increase in δ_h strengthens h's bargaining power. In particular, h becomes willing to reject more of l's offers, and this forces l to increase p_l^*. This weakens l's bargaining position, allowing h to increase its own offer p_h^*. This effect operates *weakly* in our model, and might be called the **Rubinstein effect**. However in the current model, one or even both offers may remain constant when δ_h increases. As will become clear, we might call this the **Corner effect**. Firstly it might be that $p_l^* = B_l(p_h^*) = p_l^m$ and h strictly prefers to accept p_l^m. A small increase in δ_h does not change this, so l still offers p_l^m and therefore h's offer is unchanged as well. Secondly it might be that $p_h^* = \hat{p}$. Even if higher δ_h causes p_l^* to increase, h either *cannot* or *does not want* to raise its own offer - so only p_l^* increases.

δ_l

If δ_l is too low, no prices are incentive compatible in the pricing supergame. We therefore restrict attention to $\delta_l \in [\bar{\delta}_l, 1)$, where $\bar{\delta}_l$ is the threshold value of δ_l such that p_l^m is the only price satisfying (5.1) and (5.2).

Proposition 44 *As δ_l increases, p_h^* first increases and then decreases. If h is impatient, $p_l^* = p_l^m, \forall \delta_l \in [\bar{\delta}_l, 1)$. If h is patient, $p_l^* = p_l^m$ for low and high values of δ_l, but for medium values of δ_l, p_l^* tracks p_h^**

The Rubinstein and Corner effects are still present. Therefore, *for a given bargaining set* Ω, an increase in δ_l will (weakly) reduce p_l^* and p_h^*. However the important point is that *the bargaining set is not given.* In particular as l becomes more patient, the set of IC prices expands and Ω becomes (weakly) larger. We call this the **incentive compatibility effect**. Whether this effect matters in practice depends upon whether incentive compatibility actually binds or not. For example, imagine that the firms bargain over $[p_l^m, p_h^m]$, and that their (equilibrium) offers lie below $\hat{p}$. By the Independence of Irrelevant Alternatives, they should make the same offers when bargaining over the smaller set $\Omega = [p_l^m, \hat{p}]$. Incentive compatibility is not binding, so an expansion of Ω will not affect p_l^* or p_h^*. Therefore as δ_l increases, only the Rubinstein and Corner effects apply, and price offers decline weakly.

However suppose that the solution when firms bargain over $[p_l^m, p_h^m]$ is above $\hat{p}$. Incentive compatibility does bind, and p_l^*, p_h^* are pushed down in comparison to what they would be in the $[p_l^m, p_h^m]$ bargaining game. An increase in δ_l pulls down the solutions to the $[p_l^m, p_h^m]$ bargaining game, but expands the bargaining set Ω. The latter incentive compatibility effect dominates, so p_l^* and p_h^* are pulled up.

How prices move is very much dependent on how we transit between the six cases, and this is determined by the cost asymmetry and δ_h. But in general the following pattern is observed. When $\delta_l \approx \bar{\delta}_l$, Ω is thin and the incentive compatibility effect dominates - so higher δ_l pulls the price offers upward. l is made weaker in bargaining: h can now get accepted some higher prices, which previously it could not have made since they were

not incentive compatible in the pricing game. Eventually l is sufficiently patient and the bargaining set is large. At this point the solution to the bargaining game on $[p_l^m, p_h^m]$ is below $\hat{p}$, so the Rubinstein effect dominates and price offers start to decrease. If h is impatient, we always have $B_l(p_h^*) = p_l^m$ and p_l^* never moves because h prefers not to reject low offers and suffer delay. If h is more patient, $B_l(p_h^*)$ will track p_h^* over an intermediate range, and therefore p_l^* will also have an inverted-U shape.

5.5.6 Relaxing Two Assumptions

So far we have imposed stationary bargaining strategies, and assumed that players make acceptable offers. We found a unique equilibrium satisfying these assumptions. We can actually state a stronger result.

Proposition 45 *The bargaining equilibrium satisfying only assumptions A1, A2 and A6 is automatically stationary and involves no delay*

This means that stationarity of strategies and no delay are properties of the unique equilibrium, rather than assumptions. A key driver of this result is the assumption that firms play Bertrand during disagreement games. In particular it gives firms an incentive to reach agreement earlier: if collusion at some price p is due to be reached in the future, both firms would be better off if collusion at p were reached earlier, since they would avoid playing Bertrand. This, coupled with the stationary structure of the bargaining game, implies that equilibrium will be stationary as well.

Within the context of the model the Bertrand assumption seems natural, and it is also borne out in reality. Levenstein [49] uses internal company documents to analyse a bromine cartel. Threats of price wars were used to coerce firms into joining the cartel. Price wars were also prevalent during cartel negotiations, and 'provided the ground rules' for any agreement. Price fell close to (but not below) marginal cost. Similar phenomena occurred in other industries, for example tea cartels (Levenstein and Suslow [50]).

5.6 Limit of Frictionless Bargaining

Up to now, we have assumed that it takes one unit of time both to change the market price, and to make a counteroffer. In practice offers and counteroffers can probably be made much more quickly than a price can be changed. We continue to assume that once the firms collude, they can only change their price every one unit of time. However each bargaining round is assumed to last $\Delta \to 0$ units of time. We again assume firms Bertrand compete until agreement is reached. It is analytically convenient to rewrite the earlier model in terms of flow payoffs. Hence, for example, l earns $\Pi_l(c_h)dt$ per small time interval dt when the firms Bertrand compete, and let $\delta_i^\Delta \equiv \exp(-r_i\Delta)$.

A price is incentive compatible in the collusive supergame if and only if it satisfies the same conditions as before - namely (5.1) and (5.2). This is because firms still only set a price every one unit of time. However:

Remark 46 *In the SPE of the bargaining game, l accepts any incentive compatible p_h s.t. $\Pi_l(p_h) \geq \delta_l^\Delta \Pi_l(p_l^*) + 2(1-\delta_l^\Delta)\Pi_l(c_h)$. h accepts any incentive compatible p_l s.t.*

$\Pi_h(p_l) \geq \delta_h^{\Delta}\Pi_h(p_h^*)$

Since it takes only Δ units of time to make a counteroffer, delay in bargaining only leads to Bertrand competition for Δ amount of time. Hence δ_l and δ_h in Proposition 40 must be replaced with δ_l^{Δ} and δ_h^{Δ}. We are interested in the case $\Delta \to 0$, when the time between offers becomes arbitrarily small:

Proposition 47 *As $\Delta \to 0$, l and h make the same (unique) offer, $p^* = \min(p^{NBS}, \hat{p})$ where p^{NBS} maximises $\left[\frac{1}{2}\Pi_l(p) - \Pi_l(c_h)\right]^{r_h}\left[\frac{1}{2}\Pi_h(p)\right]^{r_l}$*

As the time between offers becomes small, l and h make an identical offer which solves a constrained Nash Bargaining problem. In the limit firms make identical offers since rejection is almost costless. Therefore if one player can get collusion at p^* accepted when he offers, the other player must make him an offer arbitrarily close to p^* or it will be rejected. p^{NBS} solves an asymmetric Nash product since agreement at price p yields flow payoffs $\frac{1}{2}\Pi_l(p)$ and $\frac{1}{2}\Pi_h(p)$ whilst l's threat point is its Bertrand profit $\Pi_l(c_h)$. The overall solution is constrained because if $p^{NBS} > \hat{p}$ then the feasible price with the highest Nash product is simply $\hat{p}$ itself.

The collusive price p^* always lies strictly between p_l^m and p_h^m provided that the bargaining set Ω is non-degenerate. Intuitively, suppose it were true that $p_l^* = p_l^m$. Then h offers $p_h^* = B_h(p_l^m) \to p_l^m$ so that l is just willing to accept; h should also be willing to accept p_l^m rather than wait Δ time and get p_h^* accepted. However around these prices, l's profit function is considerably flatter than h's. Hence whilst l accepts p_h^* (by defini-

tion), h rejects p_l^m - and therefore a contradiction since in equilibrium offers are always accepted.[13]

Comparative statics are also very intuitive. When r_h falls, firm h has greater bargaining power and therefore p^* increases until it hits $\hat{p}$, after which it stays constant. Collusive price also follows an inverted-U as r_l is decreased, with the intuition being exactly as in the model with alternative offers (except that here the comparative statics are much smoother). When c_l increases, l's preferences are shifted towards higher prices and its Bertrand (threat) profits falls. It is therefore natural that the collusive price is increasing in c_l. Comparative statics in c_h are ambiguous, but price follows an inverted-U if we assume demand to be linear. Firstly the top of the bargaining set is inverted-U in c_h. In particular when c_h is small, $\hat{p} = p_h^m$ and therefore increases in c_h; when c_h is larger, $\hat{p} < p_h^m$ and $\hat{p}$ is decreasing in c_h since l's Bertrand payoff (and therefore incentives to cheat on an agreement) is increasing. Secondly - assuming linear demand - p^{NBS} also follows an inverted-U. An increase in c_h has an ambiguous effect since it shifts h's preferences towards higher prices (raises p^{NBS}) but also improves l's threat point (reduces p^{NBS}). With linear demand the first effect dominates for small c_h but the latter dominates when the cost asymmetry becomes large. Since these comparative statics are less novel and interesting compared to those in the alternating offers game, I omit further details.[14]

[13] Mathematically, $\Pi_l'(p)\Pi_h(p)r_h + \Pi_h'(p)[\Pi_l(p) - 2\Pi_l(c_h)]r_l$ is the derivative of the Nash Product, and it cannot be zero at $p = p_l^m$ or $p = p_h^m$.

[14] Harrington uses symmetric Nash bargaining, with firms offering a collusive price and market share. He has the same basic comparative statics in c_l and c_h as us. He uses linear demand as well to gain tractability.

5.7 Entry into Collusive Markets

Entry into collusive markets can have unexpected consequences when firms are asymmetric. In particular entry may make collusion *easier* to sustain, and even the arrival of a very efficient entrant can lead to consumers paying more. I assume throughout that entry is costless.

5.7.1 Entry into a Monopolistic Market

Low-cost entry is good for consumers and efficiency, whilst high-cost entry has ambiguous effects.

Low-cost entry Suppose that h is the incumbent and l enters the market. When $c_h - c_l$ is sufficiently large, collusion breaks down because the entrant has large incentives to cheat on any agreement. Bertrand competition ensues, price falls, and the efficient firm produces everything. When the entrant's cost is smaller, collusion is incentive compatible but half of all output is produced more efficiently, and price falls if for example $\Delta \to 0$ (or if h is relatively impatient and/or Ω is thin).

High-cost entry Suppose l is now the incumbent and h enters the market. If $c_h \geq p_l^m$ then nothing changes. As c_h falls below p_l^m, initially l prefers to undercut the entrant (and get the whole market) rather than collude (and only get half). In this range, lowering c_h depresses the market price and increases total surplus. However if $\delta_l, \delta_f > \frac{1}{2}$, eventually c_h is sufficiently low that collusion at p_l^m (and possibly some higher prices) is incentive compatible. High-cost entry in this region is *bad* for efficiency because now

the entrant becomes active, so half of production is undertaken by a less efficient firm, whilst the collusive price probably rises (at least at first).

5.7.2 Entry of a Third Firm

Suppose that in addition to l and h, there is an entrant e with marginal cost $c_e \geq c_l$ (but again no entry cost). In the *pricing supergame*, firms simultaneously set a price every 1 unit of time, and demand is split equally between the firms with the lowest price. We look for trigger strategies that support equilibria in which a firm sets the same price in each period. In the preceding *bargaining game* counteroffers can be made every $\Delta \to 0$ units of time, and we look for equilibria with no delay and stationary strategies. All firms have a common discount rate r. For notational convenience, let p_2^* be the equilibrium offer from the duopoly model studied earlier.

Entry May Facilitate Collusion

In this section I completely abstract from any bargaining process, and demonstrate that a (relatively low-cost) entrant may make collusion *easier* to sustain, since it makes l and h less asymmetric.

We know from earlier that collusion between l and h is incentive compatible if and only if

$$\Pi_l(p_l^m) \geq 2(1-\delta)\Pi_l(p_l^m) + 2\delta\Pi_l(c_h) \tag{5.7}$$

because p_l^m is the easiest price to enforce in the supergame (since it reduces l's incentives

to cheat). Similarly collusion between all three firms is incentive compatible only if

$$\Pi_l(p_l^m) \geq 3(1-\delta)\Pi_l(p_l^m) + 3\delta\Pi_l(\min\{c_h, c_e\}) \tag{5.8}$$

because Bertrand ensues following any deviation, and the entrant may have a lower marginal cost than h. However (5.8) is not sufficient to get three-firm collusion: we also need to check that the two cheapest firms do not wish to undercut the most expensive firm and exclude it from collusion.[15]

Lemma 48 *Suppose $c_e = c_l$. Then there exists δ such that collusion at p_l^m is impossible with two firms, but possible with three, provided that*

$$\frac{1}{4}\Pi_l(p_l^m) < \Pi_l(c_h) < \frac{2}{3}\Pi_l(p_l^m)$$

Proof: (5.8) holds if and only if $\delta \geq \frac{2}{3}$, and l and e prefer to collude with h provided $\frac{\Pi_l(p_l^m)}{3} > \frac{\Pi_l(c_h)}{2}$ hence we need $\Pi_l(c_h) < \frac{2}{3}\Pi_l(p_l^m)$. (5.7) cannot hold if $\Pi_l(c_h) > \frac{1}{2}\Pi_l(p_l^m)$ since the required δ would exceed 1. Even if $\Pi_l(c_h) \leq \frac{1}{2}\Pi_l(p_l^m)$, the critical δ that solves (5.7) exceeds $\frac{2}{3}$ provided $\Pi_l(c_h) > \frac{1}{4}\Pi_l(p_l^m)$. ■

The standard view is that entry undermines collusion because for any price, the market must be shared amongst more participants, giving greater incentives to deviate. However in this model, a low-cost entrant wipes out l's Bertrand profits, and this makes

[15] In theory l might prefer colluding at a price $\min\{c_h, c_e\}$ to colluding as a trio at p_l^m. But if it had this preference, then it would clearly deviate on collusion at p_l^m - and so (5.8) is sufficient to rule this possibility out.

it more painful for l to cheat on an agreement. Clearly entry can only facilitate collusion if the initial asymmetry between l and h is large - in particular if $\frac{1}{4}\Pi_l\left(p_l^m\right) < \Pi_l(c_h)$. However c_h cannot be too large otherwise l and e would prefer to undercut h and exclude it from any agreement - hence we also require $\Pi_l(c_h) < \frac{2}{3}\Pi_l\left(p_l^m\right)$. Figure 5-1 illustrates this point with an example where $c_l = c_e = 0$ and $D\left(p\right) = 1-p$. The red line is the critical discount factor needed to sustain collusion between l and h: it slopes up because as c_h increases, l's incentives to cheat also increase. When $\Pi_l\left(p_l^m\right) < 2\Pi_l\left(c_h\right)$ then collusion is never possible. The blue line shows the critical discount factor when l, e and h collude together at p_l^m: any $\delta \geq \frac{2}{3}$ is sufficient provided that $c_h \leq \frac{3-\sqrt{3}}{6} = \Pi_l^{-1}\left(\frac{2}{3}\Pi_l\left(p_l^m\right)\right)$. For $c_h > \frac{3-\sqrt{3}}{6}$ both l and e prefer to undercut h and exclude it from any agreement. The critical discount factors cross at $c_h = \frac{2-\sqrt{3}}{4} = \Pi_l^{-1}\left(\frac{1}{4}\Pi_l\left(p_l^m\right)\right)$. Therefore provided that $c_h$is in the (approximate) interval of $(0.07, 0.21)$, low-cost entry improves the chances of collusion - and often for a large range of discount factors.

In a more general setting marginal costs might be drawn from distributions, and Figure 5-2 attempts to capture this. Demand is again linear with $D\left(p\right) = 1 - p$, and to simplify matters $c_l = 0$. Entry has ambiguous effects: on the one hand l receives a lower market share and therefore collusion is less attractive. But on the other hand if $c_e < c_h$ then l's Bertrand payoff is lower, so cheating on an agreement is less attractive. When c_h is low relative to c_e (Region A in the diagram), the first effect dominates and entry makes collusion harder to sustain. When instead c_e is low relative to c_h (Region C) then *entry makes collusion easier to sustain* since it reduces l's Bertrand profits and

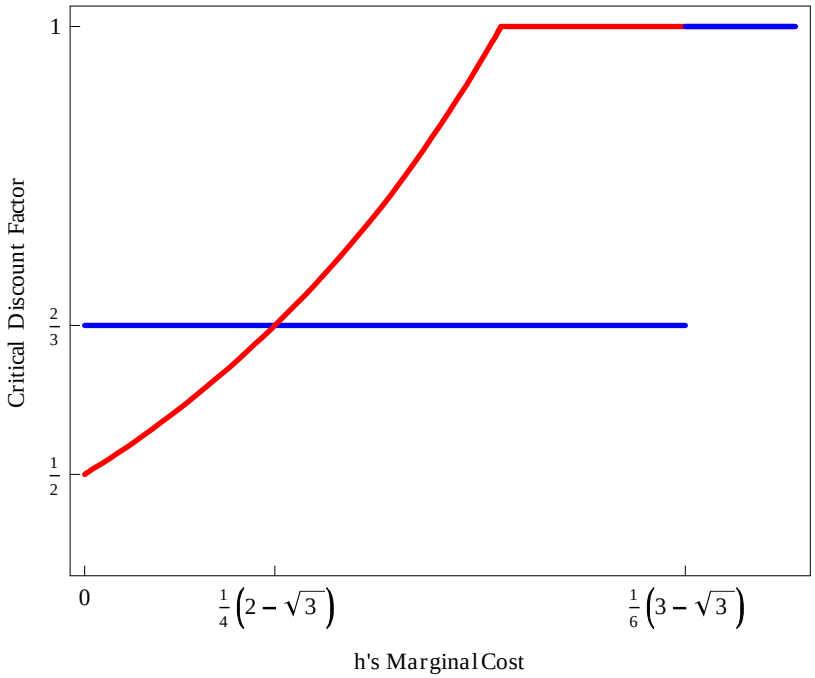


Figure 5-1: A low-cost entrant makes collusion more sustainable

therefore makes it costlier for l to cheat. When c_h and c_e are both high (Region B), l never finds it optimal to collude regardless of who else is in the market. Region D is most interesting: 3-firm collusion is incentive compatible, but l and e have different preferences. In particular l would rather exclude h whilst e would rather not. Collusion at p_l^m is therefore possible, but multiple bargaining equilibria could arise.[16]

It is clear from Figure 5-2 that if c_e and c_h are drawn randomly from the same distribution, on average entry probably makes collusion harder to sustain. Nevertheless when the entrant is reasonably efficient, it is possible to find many c_h that make collusion easier. This therefore raises the interesting possibility that entry - in particular entry of

[16] Rhodes [60] argues informally that (in a related model) the bargaining set is disjoint (includes c_h and some prices above p_l^m) and that as $\Delta \to 0$, collusion at c_h and some higher price might be possible outcomes of bargaining.

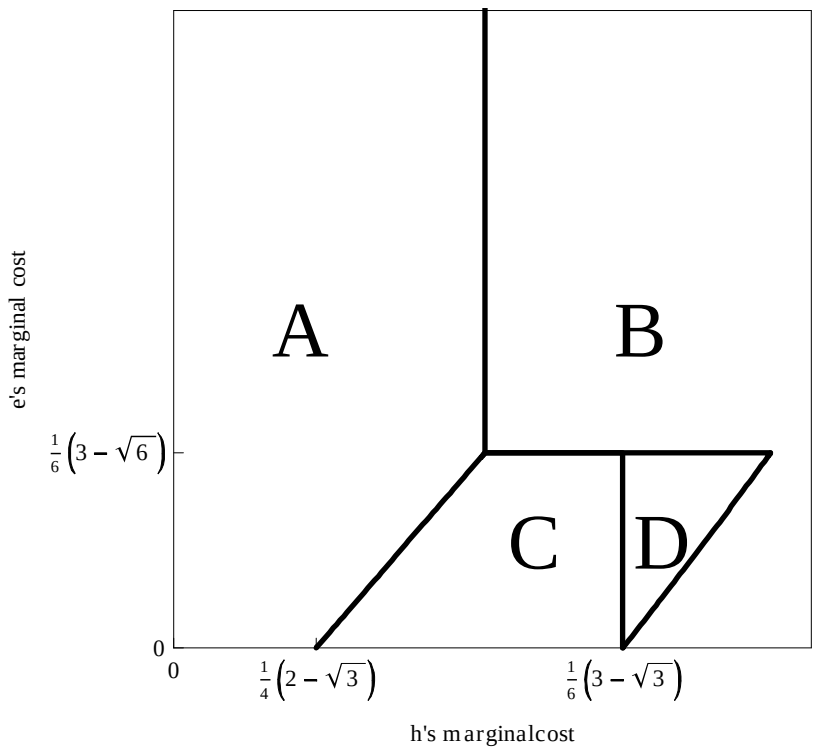


Figure 5-2: Relatively low-cost entry facilitates collusion

efficient producers - may lead to increases in market price. This is a subject we explore in more detail in the following section.

Entry and Bargaining

In this Section we assume that collusion is incentive-compatible both pre and post-entry. It is shown that a low-cost entrant may actually *raise* the collusive price, because the entrant undermines the original low-cost firm's ability to bully the high-cost firm into accepting a low price. In particular, the entrant has marginal cost $c_e \in [c_l, c_h]$ and bargains afresh with the two incumbents; to simply matters we assume that a collusive price is chosen via Nash Bargaining. Asymmetries are assumed to be small enough that

all three firms prefer to collude together. Then each firm gets one third of the market at any collusive price. Threat/disagreement payoffs are given by Bertrand profit. In Bertrand competition, l undercuts e's marginal cost and takes the entire market, so l earns $\Pi_l(c_e)$ and the other firms earn 0. Incentive compatibility is similar to earlier models. Thus the collusive price satisfies

$$\max_p \left[\frac{1}{3}\Pi_l(p) - \Pi_l(c_e)\right] \frac{1}{3}\Pi_e(p)\frac{1}{3}\Pi_h(p) \text{ s.t. incentive compatibility} \tag{5.9}$$

The Nash bargaining model acts to aggregate the preferences of the three firms. Therefore one might think that the sudden addition of a low-cost entrant would reduce the collusive price, because aggregate preferences shift towards lower prices. Similarly, high-cost entry should increase the collusive price. Provided the cost asymmetry between l and h is small, this conjecture can be shown to be broadly correct.[17] However of more interest is the following

Proposition 49 *If the cost asymmetry between l and h is large, low-cost entry* ***increases*** *the collusive price*

To understand this result, recall the two-firm bargaining model when we studied comparative statics in c_h. Higher c_h had two effects. First it pushed price up, because h's preferences switched to higher collusive prices. Second it pulled price down, because

[17] If $c_e = c_l$, collusive price falls, and if $c_e = c_h$ the collusive price would typically increase (incentive compatibility may offset this, although if asymmetries are low, incentive compatibility is less likely to bind).

l's Bertrand profit (and therefore its threat payoff) increased. Suppose that we took the two-firm model, kept h's preferences fixed, but started varying l's threat point. Clearly if we reduced this threat point, l's disagreement payoff would decline. l would then be much less able to bully h into accepting a lower price. In addition l's punishment payoff (if it deviated from collusion) would decline, more prices would become incentive compatible in the supergame, so the bargaining set might expand upwards. Therefore reducing l's Bertrand payoff would weaken l and force up the collusive price. This is exactly what a low-cost entrant does. If $c_e = c_l$, l's Bertrand profit has been reduced to zero, which ceteris paribus weakens l and pushes the collusive price up. There is, however, a second and opposing effect. The low-cost entrant itself prefers low prices, and since the Nash bargain aggregates up preferences, this will put downward pressure on the solution. Nevertheless, if the asymmetry between l and h (and therefore l's threat point in the two-firm model) is large, the collusive price rises upon entry. Intuitively the low-cost firm is usually able to use its large Bertrand profit to bully the high-cost firm into taking a low price. The entrant makes the non-collusive market so competitive, that the low-cost firm loses this ability.

Whilst high cost asymmetry between l and h is necessary for the above result, it is not actually sufficient. In the Appendix we show that *conditional on all firms colluding as a trio*, $4\Pi_l(c_h) > \Pi_l(p_l^m)$ is sufficient for the collusive price to increase. If c_h is large, this will certainly hold. However if c_h is too large, l and e might prefer to collude as a duo at c_h, rather than as a trio at some higher price. The collusive price will never be

above p_h^m, so if l and e prefer collusion at p_h^m to collusion at c_h, then in equilibrium firms will definitely collude as a trio. This requires that $\frac{1}{3}\Pi_l(p_h^m) > \frac{1}{2}\Pi_l(c_h)$ i.e. that c_h is not too large (as an extreme example, as $c_h \to c_l$, it certainly holds). Therefore in practice c_h must be large, but not too large. The following demonstrates in a linear demand model that these conditions can often be satisfied.

Example: Let $c_l = 0$, $c_h = c$, and $D(p) = 1 - bp$. Then $\Pi_l(p) = p(1 - bp)$ so $p_l^m = 1/(2b)$ and $\Pi_l(p_l^m) = 1/(4b)$. Also $p_h^m = (1 + bc)/(2b)$ so $\Pi_l(p_h^m) = \left[\frac{1+bc}{2b}\right]\left[1 - \frac{1+bc}{2}\right] = \left[\frac{(1+bc)(1-bc)}{4b}\right]$. Also $\Pi_l(c_h) = c(1 - bc)$. Then $4\Pi_l(c_h) > \Pi_l(p_l^m)$ and $\frac{1}{3}\Pi_l(p_h^m) > \frac{1}{2}\Pi_l(c_h)$ require respectively

$$16bc(1 - bc) > 1 \tag{5.10}$$

$$(1 + bc)(1 - bc) > 6bc(1 - bc) \tag{5.11}$$

We can rearrange Equation 5.11 as $(1 - bc)[1 - 5bc] > 0$. Equation 5.10 tells us that $(1 - bc) > 0$ [since $b, c > 0$] therefore Equation 5.11 tells us $[1 - 5bc] > 0$ therefore we require $bc < \frac{1}{5}$. Solving $16bc(1 - bc) > 1$ tells us that $bc > \frac{1-\sqrt{\frac{3}{4}}}{2}$. Therefore for any b, c, provided $bc \in \left(\frac{1-\sqrt{\frac{3}{4}}}{2}, \frac{1}{5}\right)$, low-cost entry will definitely raise the collusive price.

Remark 50 *If the cost asymmetry between l and h is large, high-cost entry* ***reduces*** *the collusive price*

We will not give a formal proof, however the reasoning is similar to that for low-cost

entry. Again l enjoys a Bertrand profit $\Pi_l(c_h)$ which helps it bargain down the collusive price. Once there are three instead of two firms, as a proportion of any collusive profit, $\Pi_l(c_h)$ is larger and therefore more valuable. l's threat point is worth more, placing it in a stronger bargaining position. Again there is an opposing second effect, because the entrant prefers higher prices. However if the asymmetry between l and h is large enough, the collusive price will be driven down. The set of incentive compatible prices is also likely to shrink with high-cost entry. l's deviation payoff (monopoly profit then Bertrand profit of $\Pi_l(c_h)$) is the same, but collusion with three firms is less valuable than with just two. Fewer prices are incentive compatible, and so the bargaining set may shrink.

Example: Let $D(p) = 1 - p$, $c_l = 0$ and $c_h \in \{0.03, 0.065, 0.09\}$ i.e. the asymmetry between l and h can be small ($c_h = 0.03$), medium ($c_h = 0.065$), or large ($c_h = 0.09$). We then vary c_e and compute how the bargained collusive price changes. For simplicity we set $r \to 0$ so that any p satisfying $\Pi_l(p) > 3\Pi_l(c_h)$ is incentive compatible, and this holds for the values of c_h we have chosen. Graphs plotting the collusive price against c_e are contained in Figures 5-3 - 5-5. The origin of each graph is at the collusive price that l and h would bargain to *if there were no entrant.* Then for any value of c_e, one can see whether the entrant increases or decreases the collusive price (depending upon whether the graph evaluated at that point is above or below the origin).

When the cost asymmetry between l and h is fairly small (Figures 5-3 and 5-4), low-cost entry lowers the collusive price and high-cost entry raises the price. However when

$c_e = 0.09$ the opposite is true. This value of c_e lies within the interval $\left(\frac{1-\sqrt{\frac{3}{4}}}{2}, \frac{1}{5}\right)$ given in an earlier example, so this is not surprising. The diagram confirms that when the entrant has low-cost, its main role is to undermine l's threat point, raising the collusive price compared with what it would be in the absence of entry. Conversely high-cost entry strengthens the relative value of l's threat point, pulling the collusive price down.

When the asymmetry between l and h is small (Figure 5-3), the collusive price is monotonically increasing in c_e. However for larger asymmetries (Figures 5-4 and 5-5) the collusive price follows an inverted-U in c_e. h's preferences are fixed, but as we increase c_e the entrant prefers higher prices, whilst l's threat point increases. For small asymmetries the former dominates so the price rises. For higher asymmetries the latter dominates, and the price decreases.[18]

In summary, this section demonstrates that encouraging low-cost entry into a collusive market can be bad for consumers. Low-cost entry weakens the lowest-cost incumbent in the *competitive* market. This can then lower the incentives of the low-cost firms to deviate from collusion, helping to enforce a collusive agreement, where previously one may not have existed. Low-cost entry can therefore push up the market price by *creating* collusion. Low-cost entry can also increase the collusive price if collusion already takes place. Hence if a government or regulator cares most about consumer surplus, encouraging entry *without tackling collusion* may be harmful.

[18] This is exactly as in our earlier two-firm model, where we saw that the collusive price followed an inverted-U in c_h.

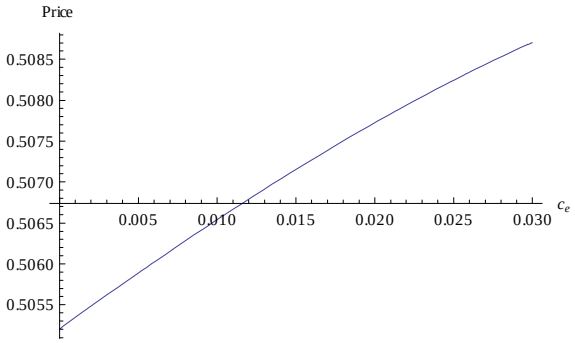


Figure 5-3: Collusion when c_h is small

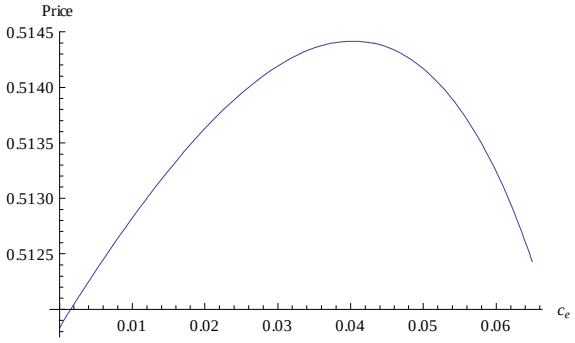


Figure 5-4: Collusion when c_h is medium

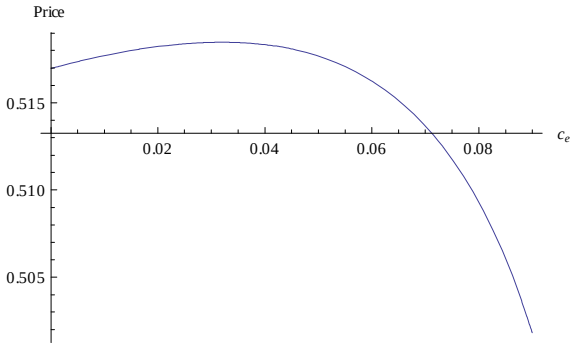


Figure 5-5: Collusion when c_h is large

5.8 Conclusion

We started off by showing that a whole continuum of prices can form part of a collusive equilibrium. An alternating offer bargaining model then enabled us to select a specific equilibrium, and gave several interesting results. In a two-firm model, the high-cost firm can have significant bargaining power, despite the low-cost firm being able to leverage its competitive position into the bargaining. Again in a two-firm model, comparative statics in the low-cost firm's patience level are non-monotonic. Greater patience can weaken the low-cost firm's bargaining strength, because higher prices become incentive compatible, enabling the high-cost firm to offer them. Our most novel results concern entry into a duopolistic market (although the principles can be extended to markets with more incumbents). We showed that entry of a low-cost firm could make the industry *more collusive*. In particular, if collusion had previously been impossible, entry might make it possible. We also showed that entry of a low-cost firm could *raise* the price in an industry that was already colluding. Both of these results stem from the fact that the low-cost entrant undermines the original low-cost firm's Bertrand profit.

5.9 Appendix

5.9.1 Properties of Best Response Functions

Best response functions have some intuitive properties which I list below (I only provide proofs for $B_l(\bar{p}_h)$).

Lemma 51 *Assume $\hat{p} > p_l^m$ and that $\bar{p}_l, \bar{p}_h \in \Omega$. Then*

- $\Pi_l(p) > 2\Pi_l(c_h) \ \forall p \in \Omega$

  Proof: Certainly $c_h < p_l^m$ or Ω is empty. So $(1-\delta_l)\Pi_l(p_l^m) + \delta_l\Pi_l(c_H) > \Pi_l(c_H)$ and using (5.1) the result follows. ■

- $B_h(\bar{p}_l) > \bar{p}_l$ *if* $\bar{p}_l < \hat{p}$; $B_l(\bar{p}_h) < \bar{p}_h$ *if* $\bar{p}_h > p_l^m$

  Proof: Take $B_h(\bar{p}_l)$. The proof is immediate if $B_h(\bar{p}_l) = \hat{p}$, so assume $B_h(\bar{p}_l) < \hat{p}$. Then $\Pi_l(B_h(\bar{p}_l)) = \delta_l\Pi_l(\bar{p}_l) + 2(1-\delta_l)\Pi_l(c_h)$, and since $\Pi_l(\bar{p}_l) > 2\Pi_l(c_h)$ and $\delta_l \in (0,1)$, $\delta_l\Pi_l(\bar{p}_l) + 2(1-\delta_l)\Pi_l(c_h) < \Pi_l(\bar{p}_l)$ and $\Pi_l(B_h(\bar{p}_l)) < \Pi_l(\bar{p}_l)$. So $B_h(\bar{p}_l) > \bar{p}_l$. The proof for $B_l(\bar{p}_h)$ is similar. ■

- *Suppose that $B_h(\bar{p}_l) = \hat{p}$. Then $\frac{\partial}{\partial \bar{p}_l} B_h(\bar{p}_l) = 0$, if (5.1) is strict then $\frac{\partial}{\partial \delta_l} B_h(\bar{p}_l) = \frac{\partial \hat{p}}{\partial \delta_l} \geq 0$, or if (5.1) binds then $\frac{\partial}{\partial \delta_l} B_h(\bar{p}_l) < 0$*

  Proof: Follows directly from (5.1). ■

- *Suppose that $B_h(\bar{p}_l) < \hat{p}$. Then $\frac{\partial}{\partial \delta_l} B_h(\bar{p}_l) < 0$, if $\bar{p}_l > p_l^m$ then $\frac{\partial}{\partial \bar{p}_l} B_h(\bar{p}_l) \in (0,1)$, or if $\bar{p}_l = p_l^m$ then $\frac{\partial}{\partial \bar{p}_l} B_h(\bar{p}_l) = 0$*

Proof: $B_h(\bar{p}_l)$ *satisfies* $\Pi_l(B_h(\bar{p}_l)) = \delta_l\Pi_l(\bar{p}_l) + 2(1-\delta_l)\Pi_l(c_h)$. δ_l *increases the righthand side so* $B_h(\bar{p}_l)$ *must fall. Also* $\frac{\partial}{\partial\bar{p}_l}B_h(\bar{p}_l) = \delta_l\left[\Pi_l'(\bar{p}_l)/\Pi_l'(B_h(\bar{p}_l))\right]$. *Then note* $\Pi_l'(p_l^m) = 0$ *and if* $\bar{p}_l > p_l^m$ *then by concavity* $0 > \Pi_l'(\bar{p}_l) > \Pi_l'(B_h(\bar{p}_l))$. ■

- *Suppose* $B_l(\bar{p}_h) = p_l^m$. *If (5.2) is strict then* $\frac{\partial}{\partial\bar{p}_h}B_l(\bar{p}_h) = \frac{\partial}{\partial\delta_h}B_l(\bar{p}_h) = 0$, *otherwise* $\frac{\partial}{\partial\bar{p}_h}B_l(\bar{p}_h) > 0$ *and* $\frac{\partial}{\partial\delta_h}B_l(\bar{p}_h) > 0$

- *Suppose* $B_l(\bar{p}_h) > p_l^m$. *Then* $\frac{\partial}{\partial\delta_h}B_l(\bar{p}_h) > 0$ *and* $\frac{\partial}{\partial\bar{p}_h}B_l(\bar{p}_h) \in (0,1)$ *provided* $\bar{p}_h < p_h^m$

Lemma 52 *There are three possible (*) combinations*

Proof: Possible combinations are {*1,*3}, {*1,*4}, {*2,*3}, {*2,*4}. But:

*1 $\Longrightarrow$ *4: *1 says $B_h(B_l(p)) = p|_{p=\hat{p}}$ and H2 says $B_h(p_l^m) < \hat{p}$. Since $B_h(B_l(p))$ increases less than one-for-one in p, $B_h(B_l(B_h(p_l^m))) > B_h(p_l^m)$ and since $B_h(\cdot)$ is weakly increasing, this implies $B_l(B_h(p_l^m)) > p_l^m$ or (*4). A similar technique shows that (*2) could imply (*3) or (*4), and that (*3) $\Longrightarrow$ (*2), and that (*4) could imply (*1) or (*2). So we can rule out {*1,*3}, rewrite {*1,*4} as *1 and rewrite {*2,*3} as *3. ■

5.9.2 General Proofs

Proof of Proposition *41*: $p_l^* = B_l(B_h(p_l^*)) = B(p_l^*)$ and from Lemma 51 $B(\cdot)$ is a weakly increasing map $\Omega \to \Omega$. By Tarski's Theorem there exist a $\check{p}_l$ such that $\check{p}_l = B(\check{p}_l)$, and since (by Lemma 51) $B(p)$ increases less than one-for-one, $\check{p}_l$ is unique and so is $B_h(\check{p}_l)$.■

Proof of Corollary *42*: If $p_l^* < \hat{p}$, then $p_h^* = B_h(p_l^*) > p_l^*$ by Lemma 51. And $p_l^* = \hat{p}$ is impossible since $B_h(p_l^*) = \hat{p}$ and by Lemma 51 $p_l^* = B_l(\hat{p}) < \hat{p}$ - a contradiction. ■

Proofs of Propositions 43 and 44: Proving Proposition 43 is similar (but simpler) than Proposition 44, so I only prove the latter. Assume that Ω is non-degenerate when $\delta_l \to 1$. We know $p_l^* = B_l(p_h^*)$ and $p_h^* = B_h(p_l^*)$ and $\frac{\partial B_l(p_h^*)}{\partial \delta_l} = 0$:

$$\frac{\partial p_l^*}{\partial \delta_l} = \frac{\partial B_l(p_h^*)}{\partial p_h} \cdot \frac{\partial p_h^*}{\partial \delta_l} \tag{5.12}$$

$$\frac{\partial p_h^*}{\partial \delta_l} = \frac{\partial B_h(p_l^*)}{\partial \delta_l} + \frac{\partial B_h(p_l^*)}{\partial p_l} \cdot \frac{\partial p_l^*}{\partial \delta_l} \tag{5.13}$$

Step 1 Begin with $\delta_l = \bar{\delta}_l$, so $\Omega = p_l^m = \hat{p}$. Case 1 holds - $B_l(p_h^*) = p_l^m$ and $B_h(p_l^*) = \hat{p}$. By Lemma 51 $B_l(p_h^*) = p_l^m \Rightarrow \frac{\partial B_l(p_h^*)}{\partial p_h} = 0$ so $\frac{\partial p_l^*}{\partial \delta_l} = 0$; also $B_h(p_l^*) = \hat{p} \implies \frac{\partial B_h(p_l^*)}{\partial \delta_l} = \frac{\partial \hat{p}}{\partial \delta_l}$ so $\frac{\partial p_h^*}{\partial \delta_l} = \frac{\partial \hat{p}}{\partial \delta_l} \geq 0$. Hence $p_l^* = p_l^m$ is static and p_h^* is equal to $\hat{p}$ (and $\hat{p}$ increases in δ_l if $\hat{p} < p_h^m$).

Step 2 δ_l is rising and $\hat{p}$ is (weakly) increasing, so eventually either $B_h(p_l^m) < \hat{p}$ or $B_l(\hat{p}) > p_l^m$.

Step 3 Suppose $B_l(\hat{p}) = p_l^m$ still, but $B_h(p_l^m) < \hat{p}$. $B_l(p_h^*) = p_l^m$ so $\frac{\partial p_l^*}{\partial \delta_l} = 0$. By Lemma 51 $B_h(p_l^m) < \hat{p} \Rightarrow \frac{\partial p_h^*}{\partial \delta_l} = \frac{\partial B_h(p_l^*)}{\partial \delta_l} < 0$. p_h^* is decreasing so $B_l(p_h^*) = p_l^m$ and $\frac{\partial p_l^*}{\partial \delta_l} = 0$ throughout. $\hat{p}$ is weakly increasing so $B_h(p_l^m) < \hat{p}$. p_h^* monotonically declines, and as $\delta_l \to 1$, $p_h^* = B_h(p_l^m) \to p_l^m$.

Step 4a Return to *Step 2* but suppose instead $B_l(\hat{p}) > p_l^m$ and $B_h(p_l^m) = \hat{p}$. Therefore

$p_h^* = \hat{p}$. If $\frac{\partial p_h^*}{\partial \delta_l} = 0$, then $\frac{\partial p_l^*}{\partial \delta_l} = 0$; if $\frac{\partial p_h^*}{\partial \delta_l} > 0$, then $\frac{\partial p_l^*}{\partial \delta_l} > 0$. Initially p_l^* and p_h^* must increase, because initially $\hat{p}$ must be increasing.[19] However there may come a point where $\hat{p}$ hits p_h^m, and then p_h^* and $p_l^* > p_l^m$ are static in δ_l.

Step 4b It remains true that $B_l(p_h^*) > p_l^m$. In *Step 4a* we started off with $B_l(\hat{p}) > p_l^m$ where $\hat{p} = p_h^*$. We then raised $\hat{p} = p_h^*$ weakly (since higher δ_l weakly increases $\hat{p}$). Hence it must continue to be true that $B_l(p_h^*) > p_l^m$.

Step 4c However eventually $B_h(p_l^*) < \hat{p}$. In *Step 4a* we started off with $B_h\left(p_l^m\right) = \hat{p}$, and therefore also $B_h\left(p_l^*\right) = \hat{p}$ since at the beginning of that step we had $p_l^* = p_l^m$. For fixed p_l^* and $p_h^* = \hat{p}$, increasing δ_l makes l more willing to reject high offers, so eventually $B_h\left(p_l^*\right) < \hat{p}$. During *Step 4a* p_l^* and $p_h^* = \hat{p}$ do (at least initially) increase. However p_l^* increases less than p_h^*. Since $p_l^* < p_h^*$ and using concavity, l's profit from p_h^* declines more than it declines from p_l^*. This reinforces the point that eventually $B_h\left(p_l^*\right) < \hat{p}$.

Step 4d When $B_l(p_h^*) > p_l^m$ and $B_h\left(p_l^*\right) < \hat{p}$, then by Lemma 51 $\frac{\partial B_l(p_h^*)}{\partial p_h}, \frac{\partial B_h(p_l^*)}{\partial p_l} \in (0,1)$ and $\frac{\partial B_h(p_l^*)}{\partial \delta_l} < 0$. Using (5.12) and (5.13), $\frac{\partial p_h^*}{\partial \delta_l} = \frac{\partial B_h(p_l^*)}{\partial \delta_l} / \left[1 - \frac{\partial B_l(p_h^*)}{\partial p_h} \cdot \frac{\partial B_h(p_l^*)}{\partial p_l}\right] < 0$ and $\frac{\partial p_l^*}{\partial \delta_l} < 0$. We are now in Case 6. Eventually p_l^* hits p_l^m, so $B_l(p_h^*) = p_l^m$ and $\frac{\partial p_l^*}{\partial \delta_l} = 0$. Still $B_h(p_l^m) < \hat{p}$ so $\frac{\partial p_h^*}{\partial \delta_l} < 0$ [*Step 3*]. Further, as $\delta_l \to 1$, $p_h^* = B_h(p_l^m) \to p_l^m$.

Conclusion When h is impatient, *Steps 1* and *3* hold. When h is patient, *Steps 1* and *4* (then finally *3* as well) hold. ■

[19] In *Step 1* $B_l(\hat{p}) = p_l^m$. h's preferences are only affected by changes in $\hat{p}$, so for us to have $B_l(\hat{p}) > p_l^m$, at that moment $\hat{p}$ must be strictly increasing.

5.9.3 Proof of Proposition 45

Following Osborne and Rubinstein [56] and Shaked and Sutton [64] we let G_i denote the set of i's bargaining SPE payoffs starting at a subgame where it makes an offer. Let $M_i = \sup G_i$ and $m_i = \inf G_i$ (we use sup and inf since G_i may be open).

Lemma 53

$$m_l = \max_{p_l \in \Omega} \frac{\frac{1}{2}\Pi_l(p_l)}{1-\delta_l} \; s.t. \; \frac{\frac{1}{2}\Pi_h(p_l)}{1-\delta_h} \geq \delta_h M_h \tag{5.14}$$

$$m_h = \max_{p_h \in \Omega} \frac{\frac{1}{2}\Pi_h(p_h)}{1-\delta_h} \; s.t. \; \frac{\frac{1}{2}\Pi_l(p_h)}{1-\delta_l} \geq \delta_l M_l + \Pi_l(c_h) \tag{5.15}$$

$$M_l = \max_{p_l \in \Omega} \frac{\frac{1}{2}\Pi_l(p_l)}{1-\delta_l} \; s.t. \; \frac{\frac{1}{2}\Pi_h(p_l)}{1-\delta_h} \geq \delta_h m_h \tag{5.16}$$

$$M_h = \max_{p_h \in \Omega} \frac{\frac{1}{2}\Pi_h(p_h)}{1-\delta_h} \; s.t. \; \frac{\frac{1}{2}\Pi_l(p_h)}{1-\delta_l} \geq \delta_l m_l + \Pi_l(c_h) \tag{5.17}$$

Proof: The proof uses standard techniques and is therefore omitted. Details are provided in Rhodes [60]. ■

Proposition 54 *Let* $\Pi_i^* = \frac{\frac{1}{2}\Pi_i(p_i^*)}{1-\delta_i}$ *where* p_i^* *is i's equilibrium offer in* σ^*.[20] *Then* in any SPE of the bargaining game satisfying A1,A2 and A6 $m_i = M_i = \Pi_i^*$, *and these are achieved by offering the same* p_i^* *as in* σ^*.

If we can prove this, we prove Proposition 45. We do this case-by-case.

Cases 1-3 **L1** says $B_l(\hat{p}) = p_l^m$ or just $\frac{\frac{1}{2}\Pi_h(p_l^m)}{1-\delta_h} \geq \frac{\frac{1}{2}\delta_h\Pi_h(\hat{p})}{1-\delta_h}$. Since $M_h \leq \frac{\frac{1}{2}\Pi_h(\hat{p})}{1-\delta_h}$ it implies $\frac{\frac{1}{2}\Pi_h(p_l^m)}{1-\delta_h} \geq \delta_h M_h$. Then by (5.14) and (5.16) l offers collusion at p_l^m and earns $m_l = M_l = \frac{1}{2}\Pi_l(p_l^m)/1-\delta_l$. Similarly **H1** implies $m_h = M_h = \frac{1}{2}\Pi_h(\hat{p})/1-\delta_h$.

[20] σ^* *is just the strategies derived on the assumptions of no delay and stationarity.*

If L1 and H1 hold, the proof is complete. If L1 and H2 hold, then $m_l = M_l = \frac{1}{2}\Pi_l(p_l^m)/1-\delta_l$ and l offers p_l^m whilst from (5.15) and (5.17), h offers a price $B_h(p_l^m)$ and earns $m_h = M_h = \frac{\frac{1}{2}\Pi_h\left(B_h(p_l^m)\right)}{1-\delta_h} = \Pi_h^*$. A similar technique is used if L2 and H1 hold.

Case 4 Clearly $m_h \leq \frac{\frac{1}{2}\Pi_h(\hat{p})}{1-\delta_h}$. If $m_h < \frac{\frac{1}{2}\Pi_h(\hat{p})}{1-\delta_h}$, then h could offer $p_h' < \hat{p}$ where $m_h = \frac{\frac{1}{2}\Pi_h(p_h')}{1-\delta_h}$ and l accept.[21] By (5.16) M_l is attained when l offers $B_l(p_h')$, so by (5.15) $m_h = \frac{\frac{1}{2}\Pi_h(B_h(B_l(p_h')))}{1-\delta_h}$. But using 51 again, *1 implies $B_h(B_l(p_h')) > p_h', \forall p_h' \in [p_l^m, \hat{p})$ so $m_h > \frac{\frac{1}{2}\Pi_h(p_h')}{1-\delta_h}$ - a contradiction. Therefore $M_h = m_h = \frac{\frac{1}{2}\Pi_h(\hat{p})}{1-\delta_h} = \Pi_h^*$ and h can only achieve this by offering $\hat{p}$ and l accepting. Then we know l offers price $B_l(\hat{p})$ and gets payoff $M_l = m_l = \frac{\frac{1}{2}\Pi_l(B_l(\hat{p}))}{1-\delta_l} = \Pi_l^*$. The proofs for **Cases 5 and 6** are similar and therefore omitted. ■

5.9.4 Proofs for Limit of Frictionless Bargaining and Entry

Proof of Proposition 47: The derivative of the unconstrained asymmetric Nash Product is proportional to $\Gamma(p)$, where

$$\Gamma(p) = \Pi_l'(p)\Pi_h(p)r_h + \Pi_h'(p)[\Pi_l(p) - 2\Pi_l(c_h)]r_l$$

Note $\Gamma(p_l^m) > 0$, $\Gamma(p_h^m) < 0$, $\Gamma(\cdot)$ is continuous and $\Gamma_p(\cdot) < 0$ so there is a unique $\check{p} \in (p_l^m, p_h^m)$ where $\Gamma(\check{p}) = 0$. Assume Ω is non-degenerate. The first steps of the proof are similar to Binmore *et al* [12].

[21] Other ways to give h a payoff m_h would involve delay and a later collusive price above p_h', so l accepts p_h'.

Step 1 $\Delta \to 0 \Rightarrow \delta_i^\Delta \to 1$, $i = l, h$. Therefore $p_l^* = B_l(p_h^*) \to p_h^* = p^*$.

Step 2 Also $B_l(\hat{p}) \to \hat{p} > p_l^m$ so (L2) holds. $B_h(p_l^m) \to p_l^m < \hat{p}$ so (H2) holds. Therefore we need only consider Cases 4-6.

Step 3 Write $\Psi(\Pi_l(p)) = \Pi_h(p)$. $\Psi(\cdot)$ is continuous and differentiable since $D(\cdot)$ has these properties. For $p \in \Omega$, $\Psi(\cdot)$ is the Pareto frontier.

Step 4 Differentiating, $\Psi'(\Pi_l(p)) = \Pi_h'(p)/\Pi_l'(p)$. $\Psi'(\cdot) \leq 0$ for $p \in \Omega$.

Step 5 Also, by definition, $\Psi'(\Pi_l(p)) \equiv \lim_{\Delta \to 0} \frac{\Psi(\Pi_l(p)) - \Psi(\delta_l^\Delta \Pi_l(p) + 2(1-\delta_l^\Delta)\Pi_l(c_h))}{\Pi_l(p) - \delta_l^\Delta \Pi_l(p) - 2(1-\delta_l^\Delta)\Pi_l(c_h)}$

Step 6a Consider Case 5. $p^* = p_l^m$. From Step 4, $\Psi'(\Pi_l(p^*)) = \Pi_h'(p^*)/\Pi_l'(p^*) \to -\infty$.

Step 6b In Case 5, $\Pi_l(p_h) = \delta_l^\Delta \Pi_l(p_l^m) + 2(1-\delta_l^\Delta)\Pi_l(c_h)$ and $\Pi_h(p_l^m) \geq \delta_h^\Delta \Pi_h(p_h)$.[22] Rewrite these as $\Pi_h(p_h) = \Psi(\Pi_l(p_h)) = \Psi(\delta_l^\Delta \Pi_l(p_l^m) + 2(1-\delta_l^\Delta)\Pi_l(c_h))$ and $\Pi_h(p_l^m) = \Psi(\Pi_l(p_l^m)) \geq \delta_h^\Delta \Pi_h(p_h)$. Combining them (and replacing p_l^m with p^*) gives $\Psi(\Pi_l(p^*)) \geq \delta_h^\Delta \Psi(\delta_l^\Delta \Pi_l(p^*) + 2(1-\delta_l^\Delta)\Pi_l(c_h))$.

Step 6c Using Steps 5, 6b, $\Psi'(\Pi_l(p^*)) \geq \lim_{\Delta \to 0} - \frac{(1-\delta_h^\Delta)}{(1-\delta_l^\Delta)} \frac{\Psi(\delta_l^\Delta \Pi_l(p^*) + 2(1-\delta_l^\Delta)\Pi_l(c_h))}{[\Pi_l(p^*) - 2\Pi_l(c_h)]}$. Alternatively, $\Psi'(\Pi_l(p^*)) \geq -\left\{\lim_{\Delta \to 0} \frac{(1-\delta_h^\Delta)}{(1-\delta_l^\Delta)}\right\} \left\{\lim_{\Delta \to 0} \frac{\Psi(\delta_l^\Delta \Pi_l(p^*) + 2(1-\delta_l^\Delta)\Pi_l(c_h))}{[\Pi_l(p^*) - 2\Pi_l(c_h)]}\right\}$. L'hôpital's rule gives $\frac{r_h}{r_l}$ for the first bracketed term.

Step 6d Combining Steps 6a,c: Case 5 requires $-\infty \geq -\frac{r_h}{r_l} \frac{\Pi_h(p^*)}{[\Pi_l(p^*) - 2\Pi_l(c_h)]}$ which is impossible since $\Pi_l(p^*) - 2\Pi_l(c_h) > 0$ and $r_l, r_h > 0$. *Conclusion* Case 5 never holds, so $p_l^* \neq p_l^m$.

Step 7a Consider Case 4. $p^* = \hat{p}$. From Step 4, $\Psi'(\Pi_l(p^*)) = \Pi_h'(\hat{p})/\Pi_l'(\hat{p})$.

Step 7b Case 4 has $\Pi_h(p_l) = \delta_h^\Delta \Pi_h(\hat{p})$, or equivalently $\Psi(\Pi_l(p_l)) = \Pi_h(p_l) = \delta_h^\Delta \Pi_h(\hat{p}) =$

[22] In Case 5, $p_h = B_h(p_l^m) < \hat{p}$ and $B_l(B_h(p_l^m)) = p_l^m$. Then use the modified acceptance rules in Remark 46, noting that h may strictly accept l's offer, but h's offer makes l just indifferent.

$\delta_h^{\Delta}\Psi(\Pi_l(\hat{p}))$. Also Case 4 has $\Pi_l(\hat{p}) \geq \delta_l^{\Delta}\Pi_l(p_l) + 2(1-\delta_l^{\Delta})\Pi_l(c_h)$. $\Psi'(\cdot) \leq 0$ so the latter implies $\Psi(\Pi_l(\hat{p})) \leq \Psi(\delta_l^{\Delta}\Pi_l(p_l) + 2(1-\delta_l^{\Delta})\Pi_l(c_h))$. Combining these gives $\Psi(\Pi_l(p_l)) \leq \delta_h^{\Delta}\Psi(\delta_l^{\Delta}\Pi_l(p_l) + 2(1-\delta_l^{\Delta})\Pi_l(c_h))$. $p_l = B_l(\hat{p}) \rightarrow \hat{p}$ from *Step 1*, so $\Psi(\Pi_l(\hat{p})) \leq \delta_h^{\Delta}\Psi(\delta_l^{\Delta}\Pi_l(\hat{p}) + 2(1-\delta_l^{\Delta})\Pi_l(c_h))$.

Step 7c Using the same tricks as in Step 6c, and combining with Step 7a, Case 4 requires $\frac{\Pi_h'(\hat{p})}{\Pi_l'(\hat{p})} \leq -\frac{r_h}{r_l}\frac{\Pi_h(\hat{p})}{\Pi_l(\hat{p})-2\Pi_l(c_h)}$. Rearranging, this requires $\Gamma(\hat{p}) \geq 0$.

Step 7d We showed earlier that there exists a unique $\check{p} \in (p_l^m, p_h^m)$ such that $\Gamma(\check{p}) = 0$, $\Gamma(p) < 0\ \forall p \in (\check{p}, p_h^m]$, $\Gamma(p) > 0\ \forall p \in [p_l^m, \check{p})$. Therefore if $\hat{p} < \check{p}$, $\Gamma(\hat{p}) > 0$ and Case 4 holds. *Conclusion* If $\hat{p} < p^{NBS}$, then Case 4 holds and $p^* = \hat{p}$.

Step 8 Consider Case 6. This Case has $\Pi_h(p_l) = \delta_h^{\Delta}\Pi_h(p_h)$ and $\Pi_l(p_h) = \delta_l^{\Delta}\Pi_l(p_l) + 2(1-\delta_l^{\Delta})\Pi_l(c_h)$.Using a similar method to above, these are equivalent to $-\frac{r_h}{r_l}\frac{\Pi_h(p^*)}{\Pi_l(p^*)-2\Pi_l(c_h)} = \frac{\Pi_h'(p^*)}{\Pi_l'(p^*)}$, or $\Gamma(p^*) = 0$. *Conclusion*: if $p^{NBS} \leq \hat{p}$ then $p^* = p^{NBS}$, otherwise Case 4 holds and $p^* = \hat{p}$. ■

Proof of Proposition 49: First show that the unconstrained Nash bargain price rises. This maximises $[(1/3)\Pi_l(p) - \Pi_l(c_e)]\,\Pi_e(p)\Pi_h(p)$. Taking logarithms and differentiating gives $\frac{\Pi_l'(p)}{\Pi_l(p)-3\Pi_l(c_e)} + \frac{\Pi_e'(p)}{\Pi_e(p)} + \frac{\Pi_h'(p)}{\Pi_h(p)} = 0$. Imposing $c_e = c_l$, the unconstrained price p_3 solves $\alpha(p) = 2\frac{\Pi_l'(p)}{\Pi_l(p)} + \frac{\Pi_h'(p)}{\Pi_h(p)} = 0$. $\alpha(p_l^m) > 0$, $\alpha(p_h^m) < 0$ and $\alpha'(p) < 0$ so there exists a unique solution. Let p_2 be the solution to the two-firm unconstrained problem. To prove $p_2 < p_3$, we must show that $\alpha(p_2) > 0$. From earlier, p_2 solves $\frac{\Pi_l'(p_2)}{\Pi_l(p_2)-2\Pi_l(c_h)} + \frac{\Pi_h'(p_2)}{\Pi_h(p_2)} = 0$. Substituting this in, we have $\alpha(p_2) = 2\frac{\Pi_l'(p_2)}{\Pi_l(p_2)} - \frac{\Pi_l'(p_2)}{\Pi_l(p_2)-2\Pi_l(c_h)}$. Then $\alpha(p_2) > 0$ requires $2\frac{\Pi_l'(p_2)}{\Pi_l(p_2)} > \frac{\Pi_l'(p_2)}{\Pi_l(p_2)-2\Pi_l(c_h)}$ or $2\Pi_l(p_2) - 4\Pi_l(c_h) < \Pi_l(p_2)$ since $\Pi_l'(p_2) < 0$. This reduces to

$4\Pi_l(c_h) > \Pi_l(p_2)$, for which a sufficient condition is $4\Pi_l(c_h) > \Pi_l(p_l^m)$.

Second show that low-cost entry expands upwards the set of IC prices. Let p' be the highest incentive compatible price with two-firms (of course it may exceed p_h^m and therefore $\hat{p}$). Using Proposition 38 p' is the highest p that solves $\Pi_l(p) = 2(1-\delta)\Pi_l(p_l^m) + 2\delta\Pi_l(c_h)$. If p'' is the highest incentive compatible price when there are three firms, it will be the highest p solving $\Pi_l(p) = 3(1-\delta)\Pi_l(p_l^m)$.[23] To show $p'' > p'$, we must prove $\Pi_l(p'') < \Pi_l(p')$. Using the above expressions this requires $3(1-\delta)\Pi_l(p_l^m) < 2(1-\delta)\Pi_l(p_l^m) + 2\delta\Pi_l(c_h)$ or $\delta > \delta^c = \Pi_l(p_l^m)/[\Pi_l(p_l^m) + 2\Pi_l(c_h)]$. (A3) guarantees that (5.1) holds for at least p_l^m, therefore $\delta \geq \delta^d = \frac{1}{2}\Pi_l(p_l^m)/[\Pi_l(p_l^m) - \Pi_l(c_h)]$. And $\delta^d > \delta^c$ provided that $4\Pi_l(c_h) > \Pi_l(p_l^m)$.

So if $4\Pi_l(c_h) > \Pi_l(p_l^m)$, then since $p_3 > p_2$ and low-cost entry (weakly) expands the bargaining set, equilibrium price must increase. ■

[23] h colludes provided $\delta \geq \frac{2}{3}$. l (and e) get $\Pi_l(p)/[3(1-\delta_l)]$ from colluding and $\Pi_l(p_l^m)$ from deviating (but no subsequent Bertrand profits).

Bibliography

[1] Anderson, S. and Renault, R. (1999): 'Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model', The RAND Journal of Economics 30(4), 719-735

[2] Arbatskaya, M. (2007): 'Ordered Search', RAND Journal of Economics 38(1), 119-126

[3] Armstrong, M. (1999): 'Price Discrimination by a Many-Product Firm', The Review of Economic Studies 66(1), 151-168

[4] Armstrong, M., Vickers, J. and Zhou, J. (2009): 'Prominence and Consumer Search', RAND Journal of Economics 40(2), 209-233

[5] Athey and Ellison (2008): 'Position Auctions with Consumer Search'

[6] Bae, H. (1987): 'A Price-Setting Supergame Between Two Heterogeneous Firms', European Economic Review 31(6), 1159-1171

[7] Bagnoli, M. and Bergstrom, T. (2005): 'Log-concave probability and its applications', Economic Theory 26(2), 445-469

[8] Bagwell, K. and Ramey, G. (1992): 'The Diamond Paradox: A Dynamic Resolution', mimeo

[9] Bain, J. (1948): 'Output Quotas in Imperfect Cartels', Quarterly Journal of Economics 62(4), 617-622

[10] Beggs, A. and Klemperer, P. (1992): 'Multiperiod competition with switching costs', Econometrica 60(3), 651–666

[11] Bils, M. (1989): 'Pricing in a Customer Market ', The Quarterly Journal of Economics, 104(4), 699-718

[12] Binmore, K., Rubinstein, A. and Wolinsky, A. (1986): 'The Nash bargaining solution in economic modelling', The RAND Journal of Economics 17(2), 176-188

[13] Bouckaert, J., Degryse, H. and Provoost, T. (2008): 'Enhancing Market Power by Reducing Switching Costs', Working Paper

[14] Burdett, K. and Judd, K. (1983): 'Equilibrium Price Dispersion', Econometrica 51(4), 955-969

[15] Cabral, L. (2008): 'Switching Costs and Equilibrium Prices', Working Paper

[16] Chen, Y. (1997): 'Paying customers to switch', Journal of Economics and Management Strategy 6(4), 877–897

[17] Chen, Y. and He, C. (2006): 'Paid Placement: Advertising and Search on the Internet'

[18] Chen, Y. and Riordan, M. (2007): 'Price and Variety in the Spokes Model', The Economic Journal 117, 897-921

[19] Collins, A. *et al* (2001): 'Below-cost legislation and retail conduct: evidence from the Republic of Ireland', British Food Journal 103(9), 607-622

[20] Diamond, P. (1971): 'A Model of Price Adjustment', Journal of Economic Theory 3, 156-168

[21] Doganoglu, T. (2004): 'Switching costs, experience goods and dynamic price competition', Working Paper

[22] Dubé, J.-P., Hitsch, G., and Rossi, P. (2009): 'Do switching costs make markets less competitive?', Journal of Marketing Research 46(4), 435-445

[23] Ellison, G. and Ellison, S. (2009): 'Search, Obfuscation, and Price Elasticities on the Internet', Econometrica 77(2), 427-452

[24] Farrell, J. (1986): 'A note on inertia in market share', Economics Letters 21(1), 73–75

[25] Farrell, J. and Klemperer, P. (2007): 'Coordination and Lock-In: Competition with Switching Costs and Network Effects', *Handbook of Industrial Organization*, Vol 3. M. Armstrong and R. Porter (eds.), North-Holland

[26] Farrell, J. and Shapiro, C. (1988): 'Dynamic competition with switching costs', RAND Journal of Economics 19(1), 123–137

[27] Fog B. (1956): 'How are Cartel Prices Determined?', The Journal of Industrial Economics 5(1), 16-23

[28] Friedman J. (1971): 'A Non-cooperative Equilibrium for Supergames', The Review of Economic Studies 38(1), 1-12

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

[30] Gehrig, T. and Stenbacka, R. (2002): 'Introductory offers in a model of strategic competition', Working Paper

[31] Gehrig, T. and Stenbacka, R. (2004): 'Differentiation-induced switching costs and poaching', Journal of Economics and Management Strategy 13(4), 635–655

[32] Ghose, A. and Yang, S. (2009): 'An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets', Management Science 55(10), 1605-1622

[33] Giulietti, M., Waddams, C. and Waterson, M. (2005): 'Consumer choice and competition policy: a study of uk energy markets', The economic journal 115, 949-968

[34] Haan, M. and Moraga-Gonzalez, J L (2009): 'Advertising for Attention in a Consumer Search Model', Discussion Paper

[35] Harrington J. (1989): 'Collusion among Asymmetric Firms. The Case of Different Discount Factors', International Journal of Industrial Organisation 7(2), 289-307

[36] Harrington, J. (1991): 'The Determination of Price and Output Quotas in a Heterogeneous Cartel', International Economic Review 32(4),767-792

[37] Hoch, S. *et al* (1995): 'Determinants of Store-Level Price Elasticity', Journal of Marketing Research, 32(1), 17-29

[38] Holmes, T. (1989): 'The Effects of Third-Degree Price Discrimination in Oligopoly', The American Economic Review 79(1), 244-250

[39] Kalyanaram, G. and Little, J. (1994): 'An Empirical Analysis of Latitutde of Price Acceptance in Consumer Package Goods', The Journal of Consumer Research 21(3), 408-418

[40] Kaufman, P. (1999): 'Rural Poor Have Less Access to Supermarkets, Large Grocery Stores', Rural Development Perspectives, 13(3), 19-26

[41] Kaul, A. and Wittink, D. (1995): 'Empirical Generalizations about the Impact of Advertising on Price Sensitivity and Price', Marketing Science 14(3), G151-G160

[42] Klemperer, P. (1987a): 'Markets with consumer switching costs', Quarterly Journal of Economics 102(2), 375–394

[43] Klemperer, P. (1987b): 'The competitiveness of markets with switching costs', RAND Journal of Economics 18(1), 138–150

[44] Klemperer, P. (1987c): 'Entry deterrence in markets with consumer switching costs', Economic Journal 97(Supplement), 99–117

[45] Klemperer, P. (1995): 'Competition when consumers have switching costs', Review of Economic Studies 62(4), 515–539

[46] Konishi, H. and Sandfort, M. (2002): 'Expanding demand through price advertisement', International Journal of Industrial Organization 20, 965-994

[47] Koulayev, S. (2009): 'Estimating demand in search markets: the case of online hotel bookings', Working Paper

[48] Lal, R. and Matutes, C. (1994): 'Retail Pricing and Advertising Strategies', The Journal of Business 67(3), 345-370

[49] Levenstein, M. (1997): 'Price Wars and the Stability of Collusion: A Study of the Pre-World War I Bromine Industry', The Journal of Industrial Economics 45(2), 117-136

[50] Levenstein, M. and Suslow, V. (2006): 'What Determines Cartel Success?', Journal of Economic Literature 44(1), 43-95

[51] MacDonald, J. (2000): 'Demand, Information, and Competition: Why do Food Prices Fall at Seasonal Demand Peaks?', The Journal of Industrial Economics 48(1), 27-45

[52] Mazumdar, T. (2005): 'Reference Price Research: Review and Propositions', Journal of Marketing 69, 84-102

[53] Milgrom, P. (1981): 'Good News and Bad News: Representation Theorems and Applications', The Bell Journal of Economics, 12(2), 380-391

[54] Milyo, J. and Waldfogel, J. (1999): 'The Effect of Price Advertising on Prices: Evidence in the Wake of 4 Liquormart', The American Economic Review 89(5), 1081-1096

[55] Muthoo, A., *Bargaining Theory with Applications* (1999), Cambridge University Press

[56] Osborne, M. and Rubinstein, A., *Bargaining and Markets* (1990), Academic Press

[57] Padilla, A.J. (1992): 'Mixed pricing in oligopoly with consumer switching costs', International Journal of Industrial Organization 10(3), 393–412

[58] Padilla, A.J. (1995): 'Revisiting dynamic duopoly with consumer switching costs', Journal of Economic Theory 67(2), 520–530

[59] Patinkin, D. (1947): 'Multiple-Plant Firms, Cartels, and Imperfect Competition', Quarterly Journal of Economics 61(2), 173-205

[60] Rhodes, A. (2007): 'Price Collusion when Firms have Different Costs', M.Phil. Thesis

[61] Rotemberg, J. and Saloner, G. (1986): 'A Supergame-Theoretic Model of Price Wars during Booms', The American Economic Review, 76(3), 390-407

[62] Rubinstein, A. (1982): 'Perfect Equilibrium in a Bargaining Model', Econometrica 50(1), 97-109

[63] Schmalensee, R. (1987): 'Competitive Advantage and Collusive Optima', International Journal of Industrial Organization 5(4), 351-367

[64] Shaked, A. and Sutton, J. (1984): 'Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model', Econometrica 52(6), 1351-1364

[65] Simester, D. (1995): 'Signalling Price Image Using Advertised Prices', Marketing Science 14(2), 166-188

[66] Sorensen, A. (2000): 'Equilibrium Price Dispersion in Retail Markets for Prescription Drugs', The Journal of Political Economy, 108(4), 833-850

[67] Stahl, D. (1989): 'Oligopolistic Pricing with Sequential Consumer Search', The American Economic Review 79(4), 700-712

[68] Stiglitz, J. (1979): 'Equilibrium in Product Markets with Imperfect Information', The American

[69] Taylor, C. (2003): 'Supplier surfing: Competition and consumer behavior in subscription markets', RAND Journal of Economics 34(2), 223–246

[70] Taylor, G. (2009): 'Competing Search Engines with Utility Maximising Consumers and Ordinary Search Results'

[71] Tirole, J., *The Theory of Industrial Organization* (1988), MIT Press

[72] To, T. (1995): 'Multiperiod competition with switching costs: An overlapping generations formulation', Journal of Industrial Economics 44(1), 81–87

[73] Varian (2007): 'Position auctions', International Journal of Industrial Organization 25, 1163-1178

[74] Vasconcelos, H. (2005): 'Tacit collusion, cost asymmetries, and mergers', The RAND Journal of Economics 36(1), 39-62

[75] Verboven, F. (1997): 'Collusive behavior with heterogeneous firms', Journal of Economic Behavior and Organization 33(1), 121-136

[76] Viard, B. (2005): 'Do Switching Costs Make Markets More or Less Competitive?: The Case of 800-Number Portability', Working Paper

[77] Viard, B. (2008): 'Do switching costs make markets more or less competitive? The case of 800-number portability', RAND Journal of Economics 38(1), 146-163

[78] Villas-Boas, M. (1999): 'Dynamic competition with customer recognition', RAND Journal of Economics 30(4), 604–631

[79] Villas-Boas, M. (2006): 'Dynamic competition with experience goods', Journal of Economics and Management Strategy 15(1), 37–66

[80] Warner, E. and Barsky, R. (1995): 'The Timing and Magnitude of Retailer Store Markdowns: Evidence from Weekends and Holidays', The Quarterly Journal of Economics, 110(2), 321-352

[81] Weitzman, M. (1979): 'Optimal Search for the Best Alternative', Econometrica 47(3), 641-654

[82] Wernerfelt, B. (1994): 'Selling Formats for Search Goods', Marketing Science 13(3), 298-309

[83] Whinston, M., *Lectures on Antitrust Economics* (2006), MIT Press

[84] White, A. (2009): 'Search Engines: Left Side Quality versus Right Side Profits'

[85] Wilson, C. (2008): 'Ordered Search and Equilibrium Obfuscation'

[86] Wilson, C. (2009): 'Market Frictions: A Unified Model of Search and Switching Costs', Working Paper

[87] Zhou, J. (2007): 'A Hotelling Model with Prominence and Non-Random Search', Unpublished Note