Three Essays in Applied Microeconomic Theory

by

Dmitry Y. Lubensky

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2011

Doctoral Committee:

Associate Professor Kai-Uwe Kühn
Professor Francine Lafontaine
Assistant Professor Stephan Lauermann
Assistant Professor Yusuf Can Masatlioglu

ACKNOWLEDGEMENTS

I am very grateful to the members of my committee – Kai-Uwe Kühn, Francine Lafontaine, Stephan Lauermann, and Yusucan Masatlioglu. They have been generous with their time and have contributed significantly to my ability to conduct research. They have also, each in their own way, provided invaluable guidance both personally and professionally. In addition, I would like to thank professors Tilman Börgers, Ying Fan, Jeremy Fox, Daisuke Nakajima, Uday Rajan, and Lones Smith for providing thoughtful feedback during the IO/Theory seminar as well as on their own time.

A special thanks goes to Mike Stevens, my co-author on the second chapter and provider of countless hours of support and feedback on chapters one and three. A similar thanks goes to Josh Cherry, Collin Raymond and Doug Smith, who along with Mike by this point know these papers better than I do. I count myself very fortunate to have had such talented and giving peers in the graduate program.

I am indebted to Amanda Berhenke and our other two roommates for the support and care they provided me, especially at times when progress was slow. Finally, I want to thank my parents Yury and Lena Lubensky and my older brother Sergei for believing in my abilities and encouraging me to find a job that I love.

'The first principle is that you must not fool yourself - and you are the easiest person to fool."
-feynman (thanks Josh)

TABLE OF CONTENTS

LIST OF FIGURES

Figure

CHAPTER I

A Model of Recommended Retail Prices

1.1 Introduction

Manufacturers routinely use non-binding recommended retail prices in markets ranging from common household goods found at the grocery store to big ticket items such as electronics, appliances, and cars. These recommendations come in a variety of forms (list prices, manufacturer suggested retail prices (MSRPs), sticker prices, etc.) and are made visible to consumers whether they shop at a brick and mortar retailer or online. There is consensus that price recommendations are closely linked to real market outcomes. This relationship has both been shown empirically (e.g. Faber and Janssen (2008)) and also implicitly assumed in the myriad studies that use recommendations as a proxy for transaction prices (e.g. Berry, Levinsohn, and Pakes (1995)). There is also anecdotal evidence that recommendations can directly affect the decisions of market participants. For example, when buying a new car consumers know not to accept a price at or above MSRP and strategic dealers seem to take this into account as they set prices.[1] However, despite the evidence *that* price recommendations effect behavior, our understanding of *how* they do so is quite limited. In part due to the fact that price recommendations are non-binding, the mechanism by which they have an impact and the motives of the manufacturer in making these recommendations are still not well understood.

In practice, most products tend to sell at or below their recommended price, hence a common explanation is that recommendations act as price ceilings. This story is compelling because how a manufacturer benefits from a price ceiling *is* well understood. Retailers with market power impose additional markups which hurt the manufacturer's sales and the manufacturer can resolve this problem with a price ceiling (Mathewson and Winter (1984)). Yet such an explanation of recom-

[1] While the majority of vehicles sell for prices strictly below MSRP there have been a few notable exceptions such as the Toyota Prius.

mendations is incomplete. Since price recommendations are non-binding, at least in name, it is not clear why a manufacturer would make a recommendation instead of just imposing a price ceiling directly. In addition, manufacturers often go to great lengths to publicize their recommendations through advertising or by printing them on product packaging. An explanation of recommendations as explicit price ceilings ignores the potential role played by consumers.

This paper provides an alternative explanation in which price recommendations directly affect consumers' search behavior. Rather than explicitly restraining retailers, price recommendations provide consumers with information. Consumers are uncertain about aggregate market conditions and when they engage in costly sequential price discovery, they do not know the distribution of retail prices. Consequently, when a consumer observes a particular price, she does not know if it represents a "good deal" or whether she should keep searching. With a price recommendation the manufacturer reveals this information and helps consumers decide between purchasing or continuing to search. Retailers anticipate consumers' reactions to the recommendation and adjust their prices accordingly. By this mechanism, non-binding price recommendations directed at consumers have a real impact on consumer and retailer behavior, and thus on market outcomes.

Price recommendations help consumers avoid the costs of learning about market conditions. But what incentive does the manufacturer have to provide this information? I show that by informing consumers and affecting search, the manufacturer faces a classic price-quantity tradeoff. In the main model consumers have either a high or a low valuation for the manufacturer's product. When a price recommendation induces consumers to reject high prices, the manufacturer restricts his ability to set a high wholesale price and cannot extract as much surplus from high valuation consumers. At the same time, when consumers reject high prices retailers are forced to reduce their markups and this increases sales to consumers with low valuations. Hence inducing more search trades off serving more low valuation consumers for extracting surplus from high valuation consumers, and which of these two effects is more important for the manufacturer depends on market conditions.

Since the manufacturer has a vested interest in how much search is undertaken, there is an issue of credibility. If the manufacturer has incentive to mislead consumers, consumers would rationally ignore recommendations and then no information can be transmitted. Furthermore, assuming this issue away by imposing an exogenous penalty for lying may be unrealistic. For example consider a book cover with one of two possible statements: "best seller" or "MSRP \$19.99". Assuming that the first statement must be truthful is reasonable since whether a book is a best seller is easily verified. But the price recommendation is not a falsifiable statement; it is not immediately obvious how a manufacturer can be accused of recommending a "false" price. In a main result, I show that credibility often does not have to be assumed and that the manufacturer can communicate market conditions to consumers using cheap talk.

The cheap talk result stems from the fact that the manufacturer's and consumers' interests can be aligned. In the main model, there is uncertainty about aggregate demand, modeled as the proportion of consumers with a high valuation. The manufacturer and retailers observe aggregate demand but consumers do not. Recalling the manufacturer's tradeoff to inducing search, he prefers less search in states when consumers are predominantly high types and more search in states when consumers are predominantly low types. For their part, consumers expect retailers to charge high prices when aggregate demand is high and low prices when aggregate demand is low and accordingly prefer to search more in the low demand state and search less in the high demand state. Since the two parties agree on whether more or less search is desirable in either state, by the logic in Crawford and Sobel (1982) the manufacturer can credibly convey information to consumers and the cheap talk result goes through.

To address how one should consider price recommendations from an antitrust perspective I present two policy experiments. First, I examine a ban on recommendations and find that in situations in which aggregate demand is sufficiently uncertain, the ban can reduce welfare and in particular the surplus of the manufacturer and consumers. I show that in equilibrium only consumers with high valuations engage in search, and the welfare result hinges on this observation. High valuation consumers, by virtue of having learned their type, believe it is more likely that aggregate demand is also high and hence they are on average overly pessimistic about the distribution of prices. As the amount of uncertainty about aggregate demand is increased, high demand consumers become more pessimistic and eventually no search can be supported. This results in a loss of sales to consumers with low demand due to higher downstream prices and a loss of sales to consumers with high demand that may exit the market prior to observing a price they would accept.

Given that price recommendations tend to look like price ceilings in practice, and that even in the context of the model presented here every recommendation maps to some upper bound on prices, a natural question is whether recommendations are just a substitute for price ceilings. I contend that this is actually not the case and that there is a distinction to be drawn between information and control. While a price ceiling gives the manufacturer power to directly influence retail prices, it does not allow the manufacturer to directly influence consumers' search behavior. I present an informal argument suggesting that there are outcomes that the manufacturer can achieve with recommendations that are not available with price ceilings alone.

While there is an extensive theoretical and empirical literature that deals with vertical relationships, an explicit treatment of the role of price recommendations has been largely lacking. Price recommendations have long been lumped in as an instrument of resale price maintenance with the mechanism behind them left unexplored. Recently though, the question of what impact recommended retail prices can have given that they are *non-binding* has been posed in two papers. Buhler and Gartner (2009) propose a repeated setting where the recommendation guides a retailer about which price to set under threat of future punishment. A different approach is taken by Puppe and Rosenkranz (2006), in which the price recommendation directly enters loss-averse consumers' utility functions as the reference point. My model provides a different mechanism from both of these papers. I depart from Buhler and Gartner by modeling recommendations as messages to consumers and in doing so argue that recommendations are more than vertical restraints. In addition, unlike Puppe and Rosenkranz, I provide an explanation where the role of recommendations is not to change consumer preferences but rather to provide information. That recommendations affect consumers' behavior is not assumed but instead emerges in equilibrium, and the effect that recommendations have is not fixed by the model but rather is determined by market conditions.

Although I focus on price recommendations in vertical markets, there is a literature on the role of non-binding list prices in markets where producers sell directly. An example is residential housing, in which studies such as Horowitz (1992) have empirically demonstrated the relationship between list and transaction prices. Still, the theoretical foundations for why this relationship should exist have been largely lacking. For instance, while Horowitz conjectures that list prices help inform consumers of a seller's reservation price, he does not show that such information can be credibly communicated in an equilibrium. By contrast, theoretical models of list prices in this setting rely on them being more than just information. For example, Gill and Thanassoulis (2010) explicitly characterize the effect of list prices on industry competition, but in doing so assume that list prices act as binding price ceilings. I propose a mechanism by which price communications are truly non-binding, and yet have a real impact by providing consumers with information that affects their search. While a vertical structure plays an important role in my results, this approach of modeling list prices as cheap talk can potentially help us understand their role in the housing and other related markets.

Lastly, this paper provides a modest methodological contribution to the literature on price search with aggregate uncertainty. Models in this literature either have sequential search but only two firms (Benabou and Gertner (1993)) or a larger number of firms and non-sequential search (Yang and Ye (2008)). The issue is tractability: sequential search with many firms potentially allows for equilibria where consumers follow non-stationary strategies. I provide a model in which search is sequential and there is a continuum of potential sellers, yet search strategies are stationary because information is communicated credibly using cheap talk. Hence, I show that in principle sequential search among a large number of sellers can be modeled tractably.

The rest of this paper proceeds with Section 2 which presents the model. Section 3 characterizes the full information equilibrium, followed by Section 4 which shows the result that cheap signaling can credibly convey information. Sections 5 and 6 discuss the two policy experiments: a ban on cheap communication and an introduction of price ceilings, respectively. Section 7 then concludes.

1.2 The Model

On the demand side there is a continuum of consumers with measure one. Each consumer demands a single unit of a good. Consumers draw their valuation for the good from a distribution in which with probability φ they draw the high value of 1 and with probability $1-\varphi$ they draw the low value $v < 1$. The parameter φ is uncertain and has two equally likely realizations φ_L and φ_H. Consumers learn only their own valuation and not the realization of φ.

The supply side consists of a monopolist manufacturer and a continuum of retailers with measure one. The manufacturer has zero production costs and is restricted to setting a uniform linear wholesale price w for all retailers. Each retailer has constant marginal costs made up of two components: the common wholesale price w and an idiosyncratic cost c. Retailers independently draw

c from a continuous and differentiable distribution $F(\cdot)$ with support on $[0, 1]$.

The game proceeds as follows. First, nature selects φ and the realization is observed by the manufacturer and retailers but not by the consumers. The manufacturer then chooses wholesale price w and signal $\sigma \in \{\sigma_L, \sigma_H\}$[2]. Next, retailers, having observed φ, w, σ and their own cost c, simultaneously set prices. These prices are then fixed for the rest of the game. Next, each consumer learns her own valuation and observes the manufacturer's signal σ and a price p from a randomly selected retailer. The consumer can purchase the good at price p and exit the market, exit the market without purchasing, or search. If she chooses to search, the consumer pays a search cost s, observes another price at a randomly selected retailer,[3] and at this point has the option to purchase the good at any of the prices she has seen so far, exit without purchasing, or search again. This process continues until every consumer has exited. There is no time discounting.

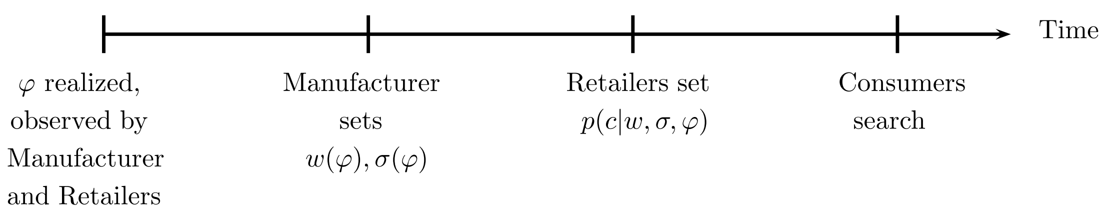

Figure 1.1: Model Timing

Strategies are $w(\varphi)$, $\sigma(\varphi)$ for the manufacturer and $p(c|w, \sigma, \varphi)$ for retailers. Define $\vec{\mathbb{P}}$ as the set of all possible price histories, so that

$$\vec{\mathbb{P}} \equiv \{\vec{p} = \{p_i\}_{i=1,\dots,n} \;\;|\;\; n \in \mathbb{I},\; p_i \in \mathbb{R}^+\}$$

Conditional on her valuation $\theta \in \{v, 1\}$, a consumer has strategy $A(\sigma, \vec{p}|\theta)$ and beliefs $\mu(\sigma, \vec{p}|\theta)$ where

$$A: \; \{\sigma_L, \sigma_H\} \times \vec{\mathbb{P}} \times \{v, 1\} \to \{\text{exit, purchase, search}\}$$

$$\mu: \; \{\sigma_L, \sigma_H\} \times \vec{\mathbb{P}} \times \{v, 1\} \to [0, 1]$$

[2]This restriction on the signal space is without loss of generality given that there are two states of nature.
[3]Specifically, I assume that when consumers search they draw every retailer with equal chance.

I use the Perfect Bayesian Equilibrium solution concept, where all strategies are mutual best responses and beliefs are formed using Bayes rule whenever possible.

1.2.1 Model Discussion

I introduce downstream retailer heterogeneity to induce price dispersion as in Reinganum (1979). On the consumer side, I depart from Reinganum's setting by having heterogenous consumers with unit demand.[4] The choice of unit demand is motivated in part by the fact that many of the goods that come with price recommendations (cars, electronics, books, etc.) are purchased one at a time as opposed to in continuous quantities. Unit demand also improves tractability while still delivering important qualitative features like downstream price dispersion and search.

The key to ensuring that the manufacturer's signals have content is to endow the manufacturer with information that consumers do not have. I have chosen aggregate demand as the source of uncertainty to reflect the fact that in many markets consumers expect prices to depend on how popular a product may be, and that sellers are more aware of this information than consumers through marketing research or other such means. In principle uncertainty can come from other sources, for instance manufacturer costs. In the conclusion, I will argue that in this scenario there is good reason to believe that the manufacturer can still inform consumers with cheap talk.

As a final note, I find that in equilibrium some retailers do not makes sales due to their high costs and the number of these retailers matters for real market outcomes. I restrict the support of retailer costs to $[0, 1]$ in order to include only retailers that can have positive gains from trade.

In the ensuing analysis, I look for an equilibrium in which price recommendations reveal the state via cheap talk. I solve for this equilibrium in two steps. In Section 1.3, I solve an auxiliary model in which there is complete information about aggregate demand φ. I then use this solution in Section 1.4 to help characterize an equilibrium of the main model where the manufacturer's signal perfectly reveals the state to consumers.

[4] Reinganum's model has homogeneous consumers with continuous demand. In either case, the key to generating price dispersion is that every retailer has different costs and faces an elastic demand function.

1.3 Equilibrium in a Full Information Setting

In this section I solve the model above but as if aggregate demand φ is common knowledge. The model is solved backwards by first characterizing the downstream equilibrium between consumers and retailers for some fixed wholesale price w and then use this solution to characterize the manufacturer's optimal choice of w.

1.3.1 Downstream Equilibrium

I show that any downstream equilibrium must be of the following form:

- low valuation consumers never search and either buy immediately or quit,
- high valuation consumers use a threshold search strategy $\bar{p}$, and
- retailers makes sales at one of two prices: either v or $\bar{p}$.

To start, in equilibrium consumers know the distribution of prices from which they sample and by a standard result in McCall (1970) they optimally follow stationary threshold strategies. Every consumer has search cost s and one implication of this is that given an equilibrium distribution of prices, if the lowest price charged is $\underline{p}$, no consumer will reject any price $p \in [\underline{p}, \underline{p} + s)$ unless it is bigger than her valuation.

Next, in any equilibrium the lowest price charged, $\underline{p}$, cannot be smaller than v. Toward a contradiction, imagine an equilibrium where $\underline{p} < v$. In such an equilibrium no consumer would reject a price $p \in [\underline{p}\ ,\ \min\{\underline{p} + s, v\})$. This implies that the retailer charging $\underline{p}$ is not maximizing profits: he can increase his price slightly and not lose any sales. Hence, no equilibrium can be supported where a price below v is charged.[5]

In turn, this implies that low valuation consumers will not collect surplus in any equilibrium and consequently never find it optimal to search after their first price observation (which is free). Thus in any equilibrium low valuation consumers either purchase at the first price if it is v or they exit.[6]

[5] This argument is adapted for this setting from its original version in Diamond (1971).

[6] Low valuation consumers are actually indifferent between buying at price v or exiting. However, for a technical reason no equilibrium can be supported when low valuation consumers reject v with positive probability. Given v is the low types' threshold, if they reject v with probability $\varepsilon > 0$, there always exists a $\delta > 0$ small enough where a retailer can charge $v - \delta$, gain the ε in sales, and make higher profits. A retailer's best response is to choose the highest price from the set of prices strictly smaller than v, but since this is an open set no such price exists. So by a closure problem an equilibrium cannot be supported for any $\varepsilon > 0$. By the same logic, high valuation consumers must accept with probability 1 at their threshold as well.

High valuation consumers may have incentive to search. Let $\bar{p}$ be their threshold and since no price is ever charged below v, it must be that $\bar{p} \geq v + s$.

Now consider the prices set by retailers. Retailers potentially face three kinds of consumers: new low valuation consumers, new high valuation consumers, and high valuation consumers that have already seen some prices but continued to search. Retailers face a step demand function: they can set a price $p \leq v$ and serve all the consumers that visit them or they can set a price $p \in (v, \bar{p}]$ and serve only the high types. A price $p > \bar{p}$ would result in no sales. Letting $\varphi\kappa$ be the number of high type searchers that visit any particular retailer, with κ later defined, the retailer's demand function is

$$q(p) = \begin{cases} 1 + \varphi\kappa & \text{if } p \leq v \\ \varphi + \varphi\kappa & \text{if } v < p \leq \bar{p} \\ 0 & \text{if } p > \bar{p} \end{cases}$$

and is illustrated in Figure 1.2.

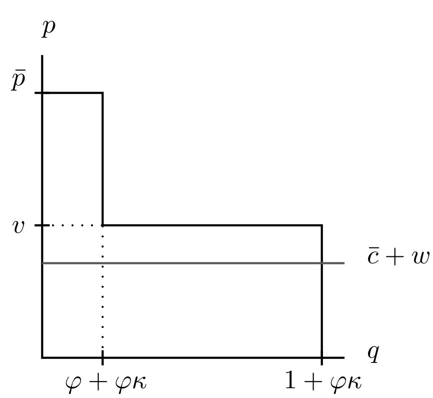

Figure 1.2: Retailer demand

Retailers choose price p to maximize their profit, given by

$$\pi(p|c) = (p - c - w)q(p)$$

Any retailer with a cost $c > \bar{p} - w$ is priced out of the market. For the purpose of exposition, I assume that all priced out retailers charge a price $p = \infty$.[7] Retailers that can afford to participate then charge either v or $\bar{p}$ depending on their cost. Define $\bar{c}$ as the cost at which a retailer is indifferent between these two prices:

$$\pi(v|\bar{c}) = \pi(\bar{p}|\bar{c}) \qquad \Leftrightarrow \qquad \bar{c} = v - w - \frac{\varphi}{1-\varphi}(\bar{p} - v)(1+\kappa) \tag{1.1}$$

If the $\bar{c}$ that solves equation (1.1) is negative, it must be that all retailers prefer to charge the higher price $\bar{p}$. Figure 1.3 illustrates an equilibrium distribution of prices when $\bar{c} > 0$.

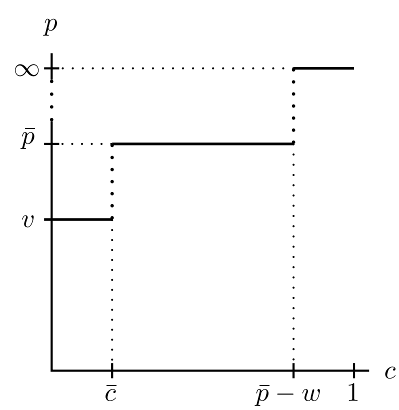

Figure 1.3: Downstream prices

To this point the structure of the equilibrium has been established: consumers follow a threshold strategy that depends on their valuations and retailers follow a threshold strategy that depends on their cost. The low types' threshold has already been shown to be their valuation v, now it remains to solve for thresholds $\bar{p}$ and $\bar{c}$.

Let $V(p)$ be the high type's value function given that p is the lowest price she has seen. By definition,

$$V(p) = \max\{0,\ 1-p,\ V^s(p)\} \tag{1.2}$$

where the consumer's options are to exit, accept p, or continue to search and receive continuation value $V^s(p) \equiv E[V(p')|p] - s$. Given the distribution of equilibrium prices summarized by $\bar{c}$, in

[7] In any equilibrium in the full information setting where the highest accepted price is $\bar{p}$, the set of prices set by priced out retailers does not affect any equilibrium outcomes.

equilibrium the high type's threshold must satisfy

$$\begin{aligned} 1-\bar{p} &= \max\Big\{0, V^s(\bar{p})\Big\} \\ &= \max\Big\{0,\ F(\bar{c})(1-v)+(1-F(\bar{c}))(1-\bar{p})-s\Big\} \end{aligned} \tag{1.3}$$

The left hand side is the value to accepting $\bar{p}$ and the right hand side gives the maximum of either the value of exiting or the continuation value of searching. When continuing to search, a consumer will pay cost s and observe price v with probability $F(\bar{c})$ or a price at least as high as $\bar{p}$ with probability $1-F(\bar{c})$. If v is observed, the consumer will accept it and obtain payoff $1-v$. If a price of $\bar{p}$ or higher is observed the continuation payoff to the consumer is $1-\bar{p}$ since she is indifferent to accepting $\bar{p}$ at that point. When $\bar{p}<1$ solves the equation above, the expression simplifies to

$$(\bar{p}-v)F(\bar{c})=s \tag{1.4}$$

This gives a natural interpretation for the threshold, with the left hand side being the expected benefit from another observation and the right hand side being the expected cost.

With $\bar{p}$ defined, it is also necessary to specify a high type's behavior when she observes a price above $\bar{p}$. Since search is costly, she may find it optimal to either continue to search or exit depending on the distribution of prices. Having observed only prices above $\bar{p}$, a consumer's continuation value to searching is

$$\begin{aligned} V^s(p\geq\bar{p}) &= F(\bar{c})(1-v)+\Big(F(\bar{p}-w)-F(\bar{c})\Big)(1-\bar{p}) \\ &\quad+\Big(1-F(\bar{p}-w)\Big)\max\Big\{0,V^s(p\geq\bar{p})\Big\}-s \end{aligned} \tag{1.5}$$

That is, a consumer can either observe and accept a price of v or $\bar{p}$ or observe a higher price and have the option to search or exit.[8] Let α be the probability with which a consumer searches conditional on rejecting. When $\bar{p}<1$ solves equation (1.3), the continuation value to searching must be strictly positive, hence $\alpha=1$. When $\bar{p}=1$, that is when high type consumers accept any price at or below their valuation, the continuation value to searching is either exactly zero or strictly negative (see

[8] Note that implicit in equation (1.5) is the fact $V^s(p\geq\bar{p})$ is constant in p. This is a standard result, stemming from the fact that having access to a price one will never accept in the future provides the same continuation value regardless of what that price actually is.

equation (1.2)). In this case equation (1.5) simplifies to

$$V^s(p \geq \bar{p} = 1) = F(\bar{c})(1 - v) - s$$

If $F(\bar{c})(1-v) - s < 0$ exiting is strictly preferred to searching so $\alpha = 0$. If $F(\bar{c})(1-v) - s = 0$, then any $\alpha \in [0,1]$ is optimal for the consumer. To summarize, α must satisfy

$$\alpha = \begin{cases} 1 & \text{if } \bar{p} < 1 \\ 0 & \text{if } \bar{p} = 1,\ F(\bar{c})(1-v) - s < 0 \end{cases} \tag{1.6}$$

Lastly, the equilibrium must specify the number of searchers $\varphi \cdot \kappa$ received by each retailer. The probability that a high type consumer rejects her first price but eventually purchases is given by

$$\Pr(\text{search and buy}) = F(\bar{p} - w) \sum_{i=1}^{\infty} \Big(\alpha(1 - F(\bar{p} - w))\Big)^i = F(\bar{p} - w)\frac{\alpha(1 - F(\bar{p} - w)}{1 - \alpha(1 - F(\bar{p} - w)}$$

Since consumers are equally likely to visit every retailer on any draw, in equilibrium all participating retailers serve the same number of searchers. Hence,

$$\kappa = \frac{\Pr(\text{search and buy})}{F(\bar{p} - w)} = \frac{\alpha(1 - F(\bar{p} - w)}{1 - \alpha(1 - F(\bar{p} - w)} \tag{1.7}$$

Note than when $\alpha = 1$ and high types continue to search until they buy, this expression reduces to $\kappa = (1 - F(\bar{p} - w))/F(\bar{p} - w)$, that is the ratio of the number of retailers that induce searchers and the number that receive searchers. When $\alpha = 0$, $\kappa = 0$ since high types either purchase at the first retailer or exit. The retailer's threshold can now be directly expressed as a function of high types' strategy $\bar{p}$ and α and is given by

$$\bar{c} = v - w - \frac{\varphi}{1 - \varphi} \frac{\bar{p} - v}{1 - \alpha(1 - F(\bar{p} - w))} \tag{1.8}$$

With consumer behavior described by equations (1.3) and (1.6) and retailers' prices described by equation (1.8), the following proposition summarizes the structure of the downstream equilibrium.

Proposition I.1. *Under full information, any downstream equilibrium is characterized by thresholds $\bar{p}$ and $\bar{c}$, and probability α that satisfy equations (1.3), (1.6), and (1.8) with strategies as follows:*

- *low valuation consumers accept any $p \leq v$, else they exit*

- *high valuation consumers accept any $p \leq \bar{p}$, else they search with probability α and exit with probability $1-\alpha$*
- *retailers set prices according to*

$$p(c) = \begin{cases} v & c \in [0, \bar{c}] \\ \bar{p} & c \in (\bar{c}, \bar{p} - w] \\ \infty & c \in (\bar{p} - w, 1] \end{cases}$$

Furthermore, a downstream equilibrium always exists.

That any equilibrium must be characterized in this way follows from the line of argument presented in preceding text. Existence can be proven by showing that equations (1.3) and (1.8) must intersect and the formal proof can be found in Appendix A.

Lastly, note that while the equilibrium must be of the form described in Proposition I.1, there may still be multiple equilibria. This is due to the complementarity between $\bar{c}$ and $\bar{p}$. A higher search threshold can lead to fewer retailers charging v which can then justify the increased search threshold. For clarity of exposition I restrict attention to the downstream equilibrium with the lowest prices.

1.3.2 Upstream Decision

With downstream behavior characterized, I now solve for the manufacturer's optimal wholesale price. It is important here to recall the information structure, in particular that while retailers observe and react to the wholesale price w, consumers only observe retail prices. Hence, the manufacturer knows that although changing the wholesale price affects the prices set by retailers (as summarized by $\bar{c}$), it would not affect the search threshold $\bar{p}$ or search probability α of the high valuation consumers.

The demand function faced by a manufacturer is given by

$$\begin{aligned} Q(w) &= \underbrace{(1-\varphi)F(\bar{c}(w))}_{\text{sales to low types}} + \underbrace{\varphi F(\bar{p}-w)(1+\kappa(w))}_{\text{sales to high types}} \\ &= (1-\varphi)F(\bar{c}(w)) + \varphi \frac{F(\bar{p}-w)}{1-\alpha(1-F(\bar{p}-w))} \end{aligned} \tag{1.9}$$

The first term is the total sales to low types and this object decreases in w. Note from equation (1.8) that

$$\frac{\partial \bar{c}}{\partial w} = -1 - \frac{\varphi}{1-\varphi}(\bar{p}-v)\frac{\alpha f(\bar{p}-w)}{1-\alpha(1-F(\bar{p}-w))} < 0$$

That fewer retailers serve low types at a higher wholesale price is a consequence of two effects: a higher wholesale price reduces markups and also increases the number of searchers by excluding more retailers. Both effects make charging the higher $\bar{p}$ relatively more attractive. This can be thought of as the double marginalization effect, which in the vertical markets literature refers to the idea that retailers use their market power to charge markups above their costs. In this setting, a retailer's market power derives from the search friction and the decision to switch to charging $\bar{p}$ rather than v reflects the use of this power.

The second term in the demand function is the total sales to high types. Here the probability α that high types continue to search plays an important role. When $\alpha = 1$, the manufacturer's demand reduces to

$$Q(w|\alpha = 1) = (1-\varphi)F(\bar{c}(w)) + \varphi \cdot \mathbb{I}(w < \bar{p}) \tag{1.10}$$

Recall that consumers do not observe the wholesale price. Rather, their only information comes from the prices they observe when searching. Given an equilibrium in which $\alpha = 1$, high type consumers never exit and continue to search until they see a price below their threshold $\bar{p}$. A manufacturer can take advantage of this behavior by increasing his wholesale price to a point at which very few retailers remain in the market. In this situation high type consumers would eventually purchase, but only after searching for a long time. Consumers cannot, from the length of their search, infer that the manufacturer has increased his wholesale price, instead they interpret a long sequence of very high prices as bad luck.

Figure 1.4 illustrates a downstream demand function $Q(w|\alpha = 1, \bar{p})$, assuming that when $w = 0$ some retailers sell to low types. The figure is meant only to show manufacturer incentives *conditional* on the strategy $(\alpha = 1,\ \bar{p})$ of high type consumers; it has not yet been shown that it represents an equilibrium.

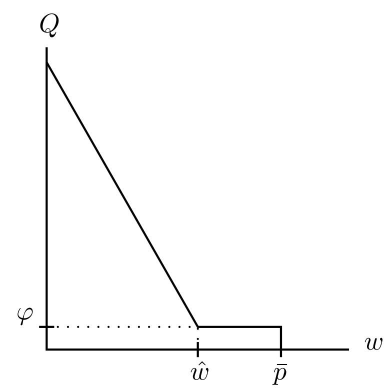

Figure 1.4: Downstream demand when $\alpha = 1$

The downward sloping part of the demand curve corresponds to wholesale prices that induce sales to both high and low type consumers. From equation (1.10), as the wholesale price w increases, sales to low types fall because $\frac{\partial \bar{c}}{\partial w} < 0$, and sales to high types remain unchanged. The demand curve kinks at the wholesale price $\hat{w}$ where $\bar{c}(\hat{w}) = 0$.[9] For higher wholesale prices, all high types continue to be served and thus the demand curve is flat. Once $w > \bar{p}$, no retailers can afford to serve the high types and the manufacturer induces no sales. Note also that as Figure 1.4 is drawn, φ is relatively small and consequently the optimal w falls in the region in which both high and low types are served. That this is actually true in equilibrium will proven later in this section.

Next consider the manufacturer's decision when high types do not always search, i.e. $\alpha < 1$. Then demand is given by

$$Q(w|\alpha < 1) = (1 - \varphi)F(\bar{c}(w)) + \varphi \frac{F(\bar{p} - w)}{1 - \alpha(1 - F(\bar{p} - w))} \tag{1.11}$$

In this case not all high type consumers are served. Every time consumers visit a retailer that is priced out they exit the market with chance $1 - \alpha$. For example, when $\alpha = 0$ the manufacturer is

[9] By equation (1.1) for any $\bar{p}$ there exists a w at which $\bar{c} = 0$.

only able to sell to high types with probability $F(\bar{p} - w)$. Figure 1.5 shows an example of such a demand function, $Q(w|\alpha = 0, \bar{p} = 1)$, with a uniform distribution of retailer costs F. Again, the same caveat as in Figure 1.4 applies in that the consumer behavior is taken as given and not shown to be an equilibrium behavior.

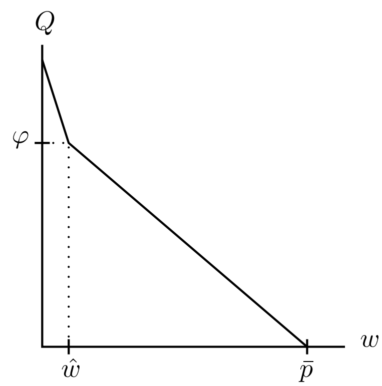

Figure 1.5: Downstream demand when $\alpha = 0$

There is still a kink in the demand curve at the wholesale price $\hat{w}$ at which $\bar{c}(\hat{w}) = 0$ and low types are no longer served. However, beyond this point increasing the wholesale price continues to result in a loss of sales. Again, as Figure 1.5 is drawn, the number of high types φ is sufficiently large so that the optimal wholesale price falls in the region in which only high types are served.

It is important for the later cheap talk result that in fact Figures 1.4 and 1.5 do depict equilibrium objects. That is, when demand φ is small enough the manufacturer serves all types in equilibrium and induces search while when demand is large enough the manufacturer serves only high types in equilibrium. This proposition is stated below with the proof relegated to Appendix B.

Proposition I.2. *The full information equilibrium is characterized by Proposition I.1 and wholesale price* $w = \arg\max\ w \cdot Q(w)$*, as given by equation (1.9). Furthermore, when aggregate demand* φ *and search cost* s *are sufficiently small, low types are served in equilibrium and high types search until they purchase. When* φ *is sufficiently large no search is induced and low types are excluded.*

1.4 Cheap Signaling

Having solved the full information model, that analysis can now be used to help characterize the equilibrium with cheap communication. I look for an equilibrium in which consumers learn the aggregate demand φ immediately upon observing the recommendation σ at the first retailer that they visit. In such an equilibrium, consumers act as if they are fully informed. But for this to be an equilibrium, when given the opportunity the manufacturer cannot have incentive to mislead consumers. That is, given that the manufacturer can induce consumers to search as if it were the high state or as if it were the low state, it needs to be verified that he will choose to signal truthfully in both states. Changing consumers' search behavior through the cheap signal amounts to a shift of the demand function faced by the manufacturer, hence showing that truthful signaling is optimal requires showing that the payoff the manufacturer obtains in the full information equilibrium is higher than the payoff he would obtain for *any* wholesale price were he to send a false signal. Cheap signaling is not always credible, instead the aim is to show that cheap signaling is possible and to highlight the conditions under which cheap signaling can arise.

First, having switched back to a game of incomplete information it is now necessary to specify consumers' beliefs. A high type consumer's belief that aggregate demand is high, $\mu(\sigma, \vec{p})$, depends on the manufacturer's signal σ and the set of prices she has observed thus far $\vec{p}$. In an equilibrium in which signals reveal the state, beliefs must be consistent with equilibrium play so that $\mu(\sigma_H, \vec{p}) = 1$ and $\mu(\sigma_L, \vec{p}) = 0$ for all price vectors $\vec{p}$ on the equilibrium path.[10] However, off the equilibrium path beliefs are unrestricted and this can lead to consumers pursuing a non-stationary search strategy.[11] To address this, I restrict attention to equilibria in which a consumer's beliefs are $\mu(\sigma_H, \cdot\,) = 1$ and $\mu(\sigma_L, \cdot\,) = 0$. This implies that when a consumer sees a price and signal combination that are inconsistent, she trusts the manufacturer's signal.

Given this restriction on beliefs, consumers act as if they are fully informed about the state. Retailers anticipate consumers' behavior as a function of recommendation σ, and since retailers expect consumers to still follow threshold strategies, they still set retail prices according to equation (1.8). I now consider the manufacturer's incentives for revealing truthfully.

Let $(\bar{p}_L, \alpha_L)$ and $(\bar{p}_H, \alpha_H)$ be the search strategies employed by consumers in the full information equilibrium in the low and high demand state, respectively. I focus on parameters such that

[10] Formally, given an equilibrium price distribution $G(p)$, any price p is on the equilibrium path if $\lim_{\varepsilon \to 0} \frac{G(p+\varepsilon)-G(p-\varepsilon)}{2\varepsilon} > 0$.

[11] As an example suppose the manufacturer sends signal σ_L and consider a price path $\vec{p} = \{p_1, p_2\}$ in which price p_1 is on the equilibrium path but p_2 is not. Since off the equilibrium path beliefs are unrestricted, suppose $\mu(\sigma_L, \{p_1, p_2\}) = 1$. Then the consumer would be using one threshold strategy $\bar{p}_L$ after observing p_1 and a different threshold strategy $\bar{p}_H$ after observing $\{p_1, p_2\}$.

under full information, search is induced in the low demand state ($\alpha_L = 1$, $\bar{p}_L < 1$) but not in the high demand state ($\alpha_H = 0$, $\bar{p}_H = 1$). By Lemmas I.6 and I.7, this requires that φ_L and s are sufficiently small and φ_H is sufficiently large. To prove the existence of a cheap talk equilibrium, I verify that conditional on being in the high demand state, the manufacturer is better off inducing $(\bar{p}_H, \alpha_H)$ than $(\bar{p}_L, \alpha_L)$ and vice versa in the low demand state. In doing this, it is important to remember that retailers are aware of the true state. While in equilibrium consumers' beliefs and hence their search behavior change in response to the manufacturer's signal, retailers know the true state and respond to the signal only in anticipation of the consumers' actions. That is, the threshold retailer $\bar{c}(w, \varphi, \bar{p}(\sigma), \alpha(\sigma))$ depends on the wholesale price w, the true state φ, and the anticipated consumer behavior $\bar{p}(\sigma)$ and $\alpha(\sigma)$.

First, suppose that aggregate demand is high. By sending signal σ_H and revealing the high state, the manufacturer earns a profit

$$\begin{aligned}\Pi^*(\sigma_H|\varphi_H) &= \max_w \; w \cdot Q(w, \sigma_H|\varphi_H) \\ &= \max_w \; w\Big((1-\varphi_H)F(\bar{c}(w, \varphi_H, 1, 0)) + \varphi_H \cdot F(1-w)\Big)\end{aligned}$$

The downstream demand is given by equation (1.11), and by assumption $\alpha_H = 0$ and $\bar{p}_H = 1$. That is, by signaling the high demand state the manufacturer induces consumers to either accept their first observed price or exit. If the manufacturer were instead to send the false signal σ_L, he would induce $\bar{p}_L < 1$ and $\alpha_L = 1$ and obtain a profit

$$\begin{aligned}\Pi^*(\sigma_L|\varphi_H) &= \max_w \; w \cdot Q(w, \sigma_L|\varphi_H) \\ &= \max_w \; w\Big((1-\varphi_H)F(\bar{c}(w, \varphi_H, \bar{p}_L, 1)) + \varphi_H \cdot \mathbb{I}(w \leq \bar{p}_L)\Big)\end{aligned}$$

The above demand function follows from equation (1.10). When sending his signal, the manufacturer thus induces one of two demand functions, as depicted in Figure 6, in which retailers' costs are distributed uniformly.

When the manufacturer signals the high state with σ_H, he faces the same demand as in Figure 1.5 in the full information equilibrium. If the manufacturer falsely signals the low state, he induces the lower search threshold $\bar{p}_L$ but also induces high types to search instead of exiting. As a result, by falsely signaling σ_L it is possible that the manufacturer can serve more consumers for wholesale prices below $\bar{p}_L$. However $\bar{p}_L$ also acts as a wholesale price ceiling on the manufacturer. The exer-

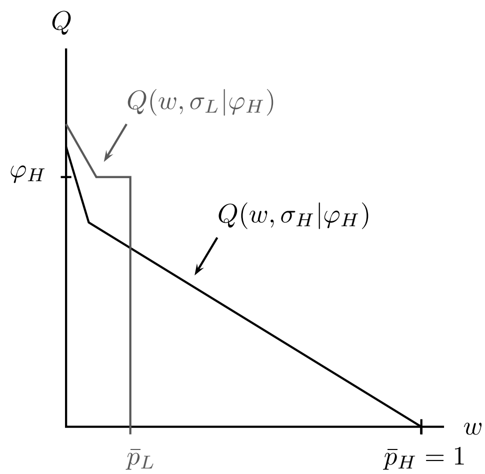

Figure 1.6: Signaling in state φ_H

cise is to show that there are parameters for which $\bar{p}_L$ is small and thus restrictive, while for those same parameters the maximized profit from truthfully signaling σ_H remains relatively high.

Lemma I.1. *Given φ_H is high enough so that no low valuation consumers are served under full information, if s, φ_L, and v are sufficiently small then the manufacturer prefers to reveal the high demand state truthfully.*

Sketch of Proof Using the fact that $\alpha_L = 1$ and $\bar{p}_L < 1$, re-arrange equation (1.3) to get

$$\bar{p}_L = v + \frac{s}{F(\bar{c}_L)} \tag{1.12}$$

Note that $\bar{c}_L$ is an equilibrium object that depends in part on the values v and s. I show in Appendix C that when s, v, and φ_L are appropriately small, the full information $\bar{p}_L$ is arbitrarily close to 0, which makes $\Pi^*(\sigma_L, \varphi_H)$ also arbitrarily close to zero. To get some intuition for this, consider the extreme case with $s = 0$. Here, all consumers will use threshold $\bar{p} = v$, and thus as v is reduced toward zero, so is the equilibrium $\bar{p}$. At the same time, fixing those values of v and s, for any φ_H the profit to truth telling $\Pi(\sigma_H, \varphi_H)$ is bounded from below by $\max_w \; w \cdot \varphi_H F(1-w) > 0$. Hence inducing $\bar{p}_L$ will reduce profits. $\square$

Next consider the low demand state φ_L. Here, the argument will be that when demand is low enough the manufacturer's optimal wholesale price will induce sales to low types regardless of which signal he sends. It will then be shown that for any wholesale price that induces sales to low types, demand is higher when the signal σ_L is sent. Figure 1.7 depicts this situation.

First it follows from equation (1.9) that regardless of search behavior α and search threshold $\bar{p} > v$, when φ is sufficiently small, that is when consumers are overwhelmingly of low type, the manufacturer will find it optimal (and feasible) to set a wholesale price low enough to induce sales to low types.

It needs to be shown that for any wholesale price that includes low types, total sales are higher with signal σ_L. Consider first sales to high types. Since $\alpha_L = 1$ every high type is eventually served while under $\alpha_H = 0$ some high types will exit prior to finding a price to which they would agree. Thus sales to high types are unambiguously higher.

Sales to low types are higher under signal σ_L if they induce a higher threshold $\bar{c}$ for retailers. The difference in $\bar{c}$ under the two signals is given by

$$\bar{c}(w, \varphi_L, \bar{p}_L, 1) - \bar{c}(w, \varphi_L, 1, 0) = \frac{\varphi_L}{1 - \varphi_L}\left(1 - v - \frac{\bar{p}_L - v}{F(\bar{p}_L - w)}\right)$$

By the proof in Appendix C, the expression $\frac{\bar{p}_L - v}{F(\bar{p}_L - w)}$ can be made arbitrarily small by choosing appropriately small values for φ_L and s. Hence, for small enough φ_L and s, it must be that $\bar{c}(w, \varphi_L, \bar{p}_L, 1) > \bar{c}(\tilde{w}, \varphi_L, 1, 0)$ and thus at every w that induces sales to low types sending the signal σ_L increases sales to low types. As a result, we obtain the following Lemma.

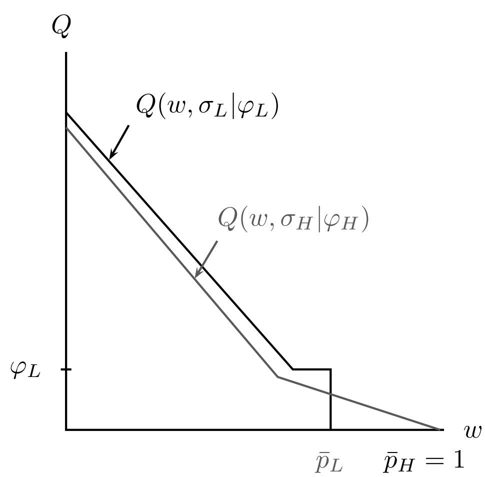

Figure 1.7: Signaling in state φ_L

Lemma I.2. *Whenever φ_H is sufficient large, if search cost s and low state aggregate demand φ_L are sufficiently small then the manufacturer prefers to reveal the low demand state truthfully.*

All that remains to check is that there exist parameter values that satisfy the conditions for truth telling in the low state and the high state. By inspection of the two preceding lemmas, as long as the low state is characterized by a small enough φ_L and v, the high state is characterized by a large enough φ_H, and the search cost s is sufficiently small, signaling can be credible in both states.

Proposition I.3. *The manufacturer can credibly communicate via cheap talk when search cost s is sufficiently small, φ_H is sufficiently high and v and φ_L are sufficiently low.*

The intuition behind why the manufacturer could credibly signal comes from understanding the impact of increasing the search threshold $\bar{p}$ on the manufacturer's demand function $Q(w)$. Increasing $\bar{p}$ worsens double-marginalization and reduces sales wholesale prices designed to serve low types. However, a higher $\bar{p}$ also means that a higher price is available at which high types can be served, and if the manufacturer intends to set a high wholesale price a higher $\bar{p}$ is beneficial.

Having established that it is possible that price recommendations can act as cheap signals that inform consumers about aggregate demand, I now consider the implications of a policy that would prohibit the manufacturer from doing so.

1.5 The Effects of a Ban on Recommendations

The ability to make price recommendations endows the manufacturer with some indirect control of downstream prices. This practice can be considered a type of vertical restraint and a natural question from an antitrust perspective is what would happen if such signaling were banned.

Without recommendations the model becomes substantially more difficult to solve as consumers' search strategies are now non-stationary and every observed price can change a consumer's beliefs about whether it is the high or low demand state. Retailers in turn would face consumers that hold heterogeneous beliefs and possibly different search thresholds. This paper will not characterize the set of equilibria with no signaling, thus a welfare effect of a ban on recommendations is not derived in general. However, a set of parameters is identified that both guarantees credible signaling and provides a tractable solution to the no-signaling equilibrium, and in these situations a clear welfare implication emerges.

In solving for the no-signaling equilibrium, the key is the belief of a consumer that has learned her type but not yet observed any prices. If a consumer learns that she has a high type, her belief is

$$\mu_0 = \frac{\varphi_H}{\varphi_H + \varphi_L}$$

with associated likelihood ratio $\lambda_0 = \frac{\varphi_H}{\varphi_L}$. Consequently, every high valuation consumer is overly pessimistic about aggregate demand (in the sense that high demand implies high prices) and every low valuation consumer is overly optimistic. Since it is only the high valuation consumers that search, consumers may search less on average than they ought to.

I focus on a setting in which λ_0 is a large number. Specifically, let φ_H be high enough so that the manufacturer finds it optimal to sell only to high types and φ_L be low enough so that the manufacturer finds it optimal to sell to all types.

Recall that $\vec{p} = \{p_1, ..., p_n\}$ represents a sequence of observed prices and $\mu(\vec{p})$ is the high valuation consumer's belief conditional on having observed $\vec{p}$. For tractability as well as to be consistent with the assumption made in the model with recommendations, I consider only equilibria in which price observations off the equilibrium path do not alter beliefs. Formally, restrict attention to equilibria with

$$\mu\big(\{p_i\}_{i=1,\dots,n}\big) = \mu\big(\{p_j\}_{j\in E}\big) \text{ where } E = \{i \mid p_i \text{ is on the equilibrium path}\} \tag{1.13}$$

In addition, I look only for equilibria in which retailers that do not make sales convey no information beyond the fact that they are priced out. Specifically, if a retailer has costs $w + c$ and in equilibrium demand at any $p \geq w + c$ is zero, then the retailer charges $p = \infty$.

I define a "no search equilibrium" to be an equilibrium of the following form. In the low demand state retailers charge either v or 1, in the high demand state retailers charge only 1. Consumers do not search after their first price draw and either accept or exit. Beliefs off the equilibrium path at all prices other than v or 1 are $\mu = \frac{\varphi_H}{\varphi_H+\varphi_L}$ by the assumption above.

Lemma I.3. *When signaling is banned, the no search equilibrium is supported whenever φ_H is sufficiently large and φ_L is sufficiently small.*

The proof can be found in Appendix D. The idea of the proof is that when a high type's prior belief $\frac{\varphi_H}{\varphi_L}$ is sufficiently high and thus pessimistic, she will not find it optimal to search and will exit if her first observation is at a retailer that is priced out.

Recall that to support a cheap talk equilibrium, it must be that φ_L, v, and s are sufficiently small and φ_H is sufficiently large. Given that the requirement to support the no search equilibrium is also that φ_H is sufficiently large and φ_L is sufficiently small, there exist parameter values where both of these equilibria are supported. I focus my analysis on this set of parameters.

Let the equilibrium with cheap signaling be described by

$$\Big((w_L^s, \bar{c}_L^s, \bar{p}_L^s, \alpha_L^s),\ (w_H^s, \bar{c}_H^s, \bar{p}_H^s, \alpha_H^s)\Big)$$

In the low state search is induced and thus $\bar{p}_L^s < 1$ and $\alpha_L^s = 1$, and low types are served so that $\bar{c}_L^s > 0$. In the high demand state sales are made only to high types and there is no search, thus $\bar{c}_H^s \leq 0$, $\bar{p}_H^s = 1$, and $\alpha_H^s = 0$. In the no search equilibrium when signaling is banned, consumers do not search and either purchase immediately or exit and manufacturer and retailer strategies are described by

$$(w_L^{ns}, \bar{c}_L^{ns}, w_H^{ns}, \bar{c}_H^{ns}; \bar{p}^{ns} = 1, \alpha^{ns} = 0)$$

In this equilibrium sales are made to low types only in the low state, thus $\bar{c}_L^{ns} > 0$ and $\bar{c}_H^{ns} \leq 0$.

Having defined the equilibria across the two regimes, let us compare the outcomes. In the high demand state, both in the signaling and no signaling cases outcomes are identical. Downstream, consumers do not search and φ_H is high enough that for any w, retailers that make sales all charge a price of 1. This implies that the manufacturer faces the same downstream demand function in both situations, and sets a wholesale price that solves

$$w_H^s = w_H^{ns} = \arg\max_w\ w \cdot \varphi_H F(1-w)$$

In the low demand state several differences emerge. First under no signaling fewer low types are served. This is because serving only high types is relatively more attractive since $\bar{p}_L^s < \bar{p}^{ns} = 1$. Also, fewer high valuation consumers are served when signaling is banned. With signaling every high valuation consumer purchases in the low state but with the ban, high valuation consumers purchase with probability $F(\bar{p}_L^{ns} - w_L^{ns}) < 1$. In addition, with no signaling those consumers that purchase do so from retailers with costs that are higher on average, since retailers with costs $c \in [\bar{p}_L^s - w_L^s, 1 - w_L^{ns}]$ can now make sales. Both reduced sales to high and low types and the increase in the average retailer costs diminish welfare. On the other hand, a ban on signaling does allow high valuation consumers that quit to save on search costs. For a high type consumer, the expected total cost of searching in a setting with signaling is given by

$$\sum_{i=1,\ldots,\infty} \Big((1 - F(\bar{p}_L^s - w_L^s) \cdot s\Big)^i F(\bar{p}_L^s - w_L^s) \;=\; F(\bar{p}_L^s - w_L^s)\frac{(1 - F(\bar{p}_L^s - w_L^s)s}{1 - (1 - F(\bar{p}_L^s - w_L^s)s}$$

Expected total search cost becomes arbitrarily small with a shrinking s. At the same time, conditional on consumers not searching both the strategies of the retailers and the manufacturer do not vary with s. Thus, as s falls the savings on search costs associated with a ban on signaling fall while the welfare losses due to fewer sales and higher average costs remain unchanged. For small enough s, the losses must dominate.

Lemma I.4. *For parameter values that support a cheap talk equilibrium and an equilibrium with no search without signaling, when search cost s is sufficiently small welfare is higher in the cheap talk equilibrium than in the no search equilibrium.*

The proof is omitted but follows along the lines of the preceding arguments.

Another way to examine the ban on recommendations is to consider its effect on the individual parties. The manufacturer is made unambiguously worse off. If the high state ensues the manufacturer is indifferent, however in the low state the outcome is as if the manufacturer successfully lies about the state to consumers, which by the cheap talk result is not in the manufacturer's interest. Low valuation consumers are equally well off as they receive zero surplus under either scenario. High valuation consumers are worse off with the ban; they receive zero surplus in the low state whereas with signaling they expect positive surplus. The effect on retailers as a group is ambiguous. Retailers that are priced out in the setting with recommendations can now make sales since $\bar{p}$ has increased to 1. But retailers with low costs make fewer sales, both due to the fact that some consumers are now served by higher cost retailers and because they receive no searchers.

1.6 A Comparison of Recommendations and Price Ceilings

Most products sell at or below their recommended price, and consequently recommendations have often been treated as binding price ceilings in the literature (Gill and Thanassoulis (2010)). While I present an alternative explanation in which recommendations communicate information, the equilibrium is still observationally equivalent to a situation in which the manufacturer sets a price ceiling

$\bar{p}_L$ and $\bar{p}_H$ in each state. At the same time, in my model the manufacturer can only induce $\bar{p}_L$ or $\bar{p}_H$ and not any other ceiling. The question is are recommendations just a poor substitute for price ceilings?

In this section I present an informal argument that the answer to this question is no, that in fact recommendations may allow a manufacturer to achieve outcomes unattainable with price ceilings alone. The essence of the argument is a distinction between information and control.

To this end, consider a setting as in the previous section in which the manufacturer is unable to communicate with consumers in any way but now can also set a binding price ceiling p^c for retailers. While consumers do not observe the ceiling[12], in equilibrium they hold beliefs about the ceiling in each state that match equilibrium strategies. What I present here is an informal argument suggestive of the fact that control is not quite enough.

Consider a set of parameters that supports a no-search equilibrium from the previous section. Without price ceilings, recall that in equilibrium retailers charge v and 1 in the low state and only 1 in the high state. As argued in Lemma I.3, a no-search equilibrium is supported because high types are convinced that it is a high demand state, in which case the returns to searching are zero. With this in mind, one may conjecture that for these parameters an equilibrium with search can be supported if the manufacturer sets a price ceiling $p^c < 1$ in the high state. Doing so would then give returns to searching in the high state, and thus even agents that are pessimistic would continue searching until they find an acceptable price.

However this equilibrium would be quite difficult to support. The issue is commitment by the manufacturer. Suppose with some price ceiling p^c the manufacturer induces the search behavior $\alpha = 1$ so that no consumer ever exits. If in the high demand state the manufacturer looks to serve only high types, then he has incentive to eat into the returns to searching by, for instance, increasing his wholesale price almost all the way up to the price ceiling. Doing so would not cost him sales since high types would continue to search until eventually buying, however it would make search no longer optimal ex-post. Since this commitment concern is commonly known, an equilibrium in which a price ceiling can deliver the requisite returns to searching in the high demand state is

[12]If p^c were public, consumer could infer from it aggregate demand, hence the ceiling would serve both as information and control and should in principle do at least as well as information alone.

difficult to obtain.

Suppose then that a manufacturer is unable to induce an equilibrium with search with price ceilings alone. If this is the case, then the comparison of outcomes with price ceilings versus with price recommendations is the same comparison as in the previous section. Namely, the manufacturer is worse off with ceilings. By communicating with consumers, the manufacturer can induce them to search if the demand is low and losing this ability can be costly.

1.7 Discussion

This paper addresses the question of how price recommendations can impact a market for consumer goods. I posit that such communication is an attempt by manufacturers to inform consumers about their returns to searching. When aggregate demand is low, the manufacturer induces more search and in doing so reduces retailer markups and increases sales to consumers with low valuations. When aggregate demand is high, the manufacturer induces less search which allows him to charge a higher wholesale price thereby extracting surplus from consumers with high valuations. I show that when search costs are low and aggregate demand is sufficiently uncertain, the manufacturer is able to credibly communicate with consumers via cheap talk.

This model draws a sharp distinction between price recommendations and resale price maintenance. Recommendations are not just an indirect method to manage double marginalization. Though the effect of recommendations is observationally similar to that of price ceilings, the mechanism is quite different — in my model the manufacturer influences outcomes by informing consumers rather than imposing constraints on retailers. And while providing information sometimes leads to reduced double marginalization, at other times it leads retailers to actually set higher prices.

While I introduce uncertainty about aggregate demand in the main model, other forms of aggregate uncertainty could potentially make cheap communication possible. For example, consider uncertainty about the manufacturer's costs. A manufacturer facing high costs would be more interested in extracting surplus from high valuation consumers, while a manufacturer facing low costs would be relatively more interested in increasing total sales. At the same time, consumers should expect higher prices in the high cost state and lower prices in the low cost state. I conjecture that when the two cost states are sufficiently different, the manufacturer can credibly communicate his costs via cheap talk.

In practice there is variation in the way actual prices relate to price recommendations. For instance, books often sell for exactly their jacket price while cars tend to sell for strictly less than MSRP. Even within the car market, how far below MSRP a car sells varies by the popularity of that vehicle, and in fact some cars sell above MSRP. Imposing that recommendations are price ceilings or some other exogenously determined restraints precludes the analysis from explaining such variation. Since in my model a recommendation is just a cheap message, I allow for the consumers' interpretation of this message to vary with market conditions and thus can accommodate the varying relationship between prices and recommendations.

The goal of this paper was to highlight a mechanism by which price recommendations serve a purely informational role and still influence market outcomes. Towards this end, the model I develop is quite stylized and thus has limitations for direct use in assessing policy. Nonetheless, the message that seems to emerge is that ceteris paribus price recommendations help consumers, allowing them to make more informed decisions and saving search costs. Hence, an antitrust policy discussion of the merits of price recommendations should keep these informational benefits in mind.

1.8 Appendix

Appendix A: Proof of Existence of the Downstream Equilibrium Under Full Information

I show that a downstream equilibrium that is characterized in Proposition I.1 must exist by arguing that equations (1.8) and (1.3) are continuous and that they must intersect.

Recall equation (1.3) which describes the high valuation consumer's threshold:

$$1 - \bar{p} = \max\{0,\ F(\bar{c})(1 - v) + (1 - F(\bar{c}))(1 - \bar{p}) - s\}$$

and let $\bar{p}^{(1.3)}(\bar{c})$ be the implicit function implied by this equation. By inspection this function is continuous. First note that $\bar{p}^{(1.3)}(\bar{c}) > v$ for all $\bar{c}$. This can be seen by contradiction. Rewrite equation (1.3) as

$$1 - \bar{p} = \max\{0,\ 1 - \bar{p} + F(\bar{c})(\bar{p} - v) - s\}$$

If $\bar{p} < v$ then the left hand side must be larger than the right hand side, hence a contradiction. Next note that because of the max operator, $\bar{p}^{(1.3)}(\bar{c}) \leq 1$ for all $\bar{c}$.

Now recall equation (1.8)

$$\bar{c} = v - w - \frac{\varphi}{1-\varphi} \frac{\bar{p} - v}{1 - \alpha(1 - F(\bar{p} - w))}$$

Recall also that $\bar{p} < 1$ implies $\alpha = 1$, else if $\bar{p} = 1$ any $\alpha \in [0, 1]$ can be used to support an equilibrium. In this sense, α helps the existence argument in that for $\bar{p} = 1$, there are many values of $\bar{c}$ that can satisfy equation (1.8) given a choice of α.

Let equation (1.8) implicitly define the function $\bar{p}^{(1.8)}(\bar{c})$. Continuity once again is obvious here. First note that $\bar{p}^{(1.8)}(v-w) = v$. Also, note that there exists a $\bar{c}$ low enough (and possibly negative) so that $\bar{p}^{(1.8)}(\bar{c}) = 1$.

Hence, $\bar{p}^{(1.8)}(v - w) < \bar{p}^{(1.3)}(v - w)$ and $\bar{p}^{(1.8)}(\bar{c}) \geq \bar{p}^{(1.3)}(\bar{c})$ for some small enough $\bar{c}$. Given that $\bar{p}^{(1.8)}(\bar{c})$ and $\bar{p}^{(1.3)}(\bar{c})$ are continuous functions, they must then intersect. $\square$

Appendix B: Proof of the Characterization of the Full Information Equilibrium

This section provides a proof to Proposition I.2, restated here.

Proposition I.2 *The full information equilibrium is characterized by Proposition I.1 and wholesale price $w = \arg\max\ w \cdot Q(w)$, as given by equation (1.9). Furthermore, when aggregate demand φ and search cost s are sufficiently small, low types are served in equilibrium and high types search until they purchase. When φ is sufficiently large no search is induced and low types are excluded.*

Proof. The proof proceeds through a series of claims.

Lemma I.5. *When φ is small enough, the optimal wholesale price chosen by the manufacturer induces sales to low types.*

Proof of Lemma I first show that when the proportion of high valuation consumers φ is small enough it is feasible for the manufacturer to induce sales to low types. I then show that it is optimal for him to do so.

By equation (1.8)

$$\begin{aligned}\bar{c} &= v - w - \frac{\varphi}{1-\varphi}\,\frac{\bar{p}-v}{1-\alpha(1-F(\bar{p}-w))}\\ &\geq v - w - \frac{\varphi}{1-\varphi}\,\frac{1-v}{F(v-w)}\end{aligned}$$

By inspection, the above expression shows that when φ is sufficiently small $\bar{c}(w = 0) > 0$. In fact, for sufficiently small φ there will be a range of wholesale prices that will induce search to low types. Equation (1.9) then shows that that as φ is decreased, the quantity sold by setting a wholesale price that serves only high types goes to zero while the quantity that can be sold with a wholesale price that includes low types is bounded strictly above zero. Hence, it is also optimal for the manufacturer to choose a wholesale price that serves low types. □

In fact, I prove a stronger statement about the low state. Specifically, if both φ and search cost s are sufficiently small, the manufacturer will optimally induce a downstream equilibrium in which $\alpha = 1$.

Lemma I.6. *When φ and search cost s are small enough, high type consumers search in equilibrium with probability $\alpha = 1$.*

Proof of Lemma Fix a high type consumer's strategy $\bar{p} < 1$ and $\alpha = 1$. By Lemma I.5 there exists a φ low enough so that the manufacturer chooses a w to induce sales to low types, and let $\bar{c}$ be the ensuing retailer threshold induced by w. In order for this to be an equilibrium, it must be that $\bar{p} < 1$ is a best response to the induced price distribution. By inspection of equation (1.3), for any $\bar{c}$ and $\bar{p} > v$, there exists an s small enough to make the equation hold. Hence $\bar{p} < 1$ can be supported in equilibrium and as a result $\alpha = 1$ is supported. □

I have shown that when φ is small, the manufacturer sets an equilibrium wholesale price that induces sales to low types. Furthermore, when the search cost s is small, consumers will search in equilibrium. Next I argue that when φ is high enough the manufacturer will set a wholesale price that excludes low types.

Lemma I.7. *When φ is large enough, for any search threshold $\bar{p} > v$ the manufacturer charges a wholesale price w that excludes low types.*

Proof of Lemma By equation (1.1)

$$\bar{c} = v - w - \frac{\varphi}{1-\varphi}(\bar{p} - v)(1+\kappa) \leq v - \frac{\varphi}{1-\varphi} \cdot s$$

The second inequality follows from the fact that $\bar{p} - v \geq s$ and $1 + \kappa \geq 1$. The above expression is negative for φ large enough, hence the manufacturer will not have the option of inducing sales to low types. □

Lastly, if there is an equilibrium in which the manufacturer only serves high types, it must be that there is no search and $\alpha = 0$. The reason is that if only high types are included, and high types follow a threshold strategy, then only a price of 1 can be supported downstream. If the only price charged is 1 then search is never worthwhile. Along with Lemmas I.5, I.6, and I.7 the proves the proposition. □

Appendix C: Proof of Arbitrarily Low Threshold $\bar{p}$ in the Full Information Equilibrium

This section proves that for any small $\bar{p} > 0$, there exist small enough φ, s, and v so that $\bar{p}$ is supported in the full information equilibrium. I will show this by first noting that a manufacturer's profits in any equilibrium are bounded away from zero. Then, I will show that as s and φ are reduced toward zero, either $\bar{p}$ approaches v or the manufacturer's profit approaches zero, which would contradict the first statement. Lastly, since $\bar{p}$ can be made arbitrarily close to v, when v is chosen to be small, $\bar{p}$ will be small as well.

First, recall that by Lemma I.6, the manufacturer chooses to set a wholesale price w to induce sales to low types, i.e. he induces $\bar{c} > 0$. It will be useful for this argument to show that as $\varphi \to 0$, the manufacturer's profit is uniformly bounded strictly above 0.

Claim I.1. *For any $\delta > 0$, there exists a low enough $\hat{\varphi}$ so that for any $\varphi < \hat{\varphi}$, the manufacturer's equilibrium profit exceeds $v/2 \cdot F(v/2) - \delta$.*

Proof of Claim: The manufacturer's equilibrium profit, given w is the optimally charged wholesale price, is given by

$$
\begin{aligned}
\Pi(w,\varphi) &= w\Big((1-\varphi)F(\bar{c}(w,\varphi)) + \varphi\cdot\mathbb{I}(w\le\bar{p})\Big) \\
&\ge w\left((1-\varphi)F\left(v-w-\frac{\varphi}{1-\varphi}\frac{\bar{p}-v}{F(\bar{p}-w)}\right)\right) && \text{(by equation (1.8))} \\
&= w\left(F(v-w) - \underbrace{\left(F\left(v-w-\frac{\varphi}{1-\varphi}\frac{\bar{p}-v}{F(\bar{p}-w)}\right) - F(v-w)\right)}_{A} - \underbrace{\varphi F\left(v-w-\frac{\varphi}{1-\varphi}\frac{\bar{p}-v}{F(\bar{p}-w)}\right)}_{B}\right)
\end{aligned}
$$

Consider a manufacturer that charges $w = v/2$. At this wholesale price, $\bar{p}-w \ge v+s-w \ge v/2+s > 0$, hence both terms A and B go to zero as φ goes to zero. For φ low enough, $v/2\cdot(A+B) < \delta$. Hence, $\Pi(w,\varphi) \ge \Pi(v/2,\varphi) \ge v/2\cdot F(v/2) - \delta$. □

Next, I prove by contradiction that as s and φ decrease, $\bar{p}-v$ approaches zero. Suppose toward a contradiction that there exists an $\varepsilon > 0$ such that there is some $\hat{s}$ with the property that for any $s < \hat{s}$, in the full information equilibrium $\bar{p}-v > \varepsilon$. By equation (1.12), this implies that $\frac{s}{F(\bar{c})} > \varepsilon$ for any $s < \hat{s}$. Recall equation (1.8):

$$
\bar{c} = v - w - \frac{\varphi}{1-\varphi}\frac{\bar{p}-v}{F(\bar{p}-w)}
$$

For small enough φ in any equilibrium $w < v$. By the hypothesis above, this implies that $\bar{p}-w \ge \bar{p}-v \ge \varepsilon$ for all $s < \hat{s}$. Then as s and φ both shrink toward zero, the right hand side of the equation above must approach $v-w$. At the same time, since by hypothesis $\frac{s}{F(\bar{c})} > \varepsilon$ for all small s, it must be that $\bar{c}$ is converging toward zero. Hence, it must be that as s and φ shrink toward zero, the equilibrium w approaches v. However, this implies that the manufacturer's equilibrium profit approaches zero which contradicts the claim above since the manufacturer's profit has an absolute lower bound strictly above zero.

Hence I have shown that as search cost s and low state demand φ_L diminish toward zero, the equilibrium search threshold $\bar{p}$ approaches the low types' valuation v. Thus, there always exists a full information equilibrium where the search threshold is arbitrarily close to zero given parameters φ, s, and v are all chosen to be sufficiently small. □

Appendix D: Proof of Existence of a No Search Equilibrium When Recommendations are Banned

When recommendations are banned, I show that an equilibrium with no search can be supported when initial belief λ_0 is sufficiently high.

Proof. In the proposed equilibrium, retailers face the step demand function

$$q(p) = \begin{cases} 1 & \text{if } p \in [0, v] \\ 1 - \varphi & \text{if } p \in (v, 1] \\ 0 & \text{if } p \in (1, \infty) \end{cases}$$

Retailers use threshold strategy $\bar{c}$ given by equation (1.1)

$$\bar{c}(w, \varphi) = v - w - \frac{\varphi}{1 - \varphi}(1 - v)$$

Given an equilibrium where consumers do not search, the manufacturer solves

$$\max_{w} \; w \cdot Q(w, \varphi) = w \cdot \Big((1 - \varphi)F(\bar{c}(w, \varphi)) + \varphi F(1 - w)\Big)$$

in each state φ, with $\bar{c}(w, \varphi)$ given above and decreasing in w. Let $w(\varphi)$ be the solution to the above optimization and note that $0 < w(\varphi) < 1$.

Claim I.2. *When φ_H is large enough, $\bar{c}(\varphi_H) \leq 0$ and when φ_L is small enough $\bar{c}(\varphi_L) > 0$ in equilibrium.*

Proof of Claim This does not follow immediately because in equation (1.1) wholesale prices are endogenous. That $\bar{c}(\varphi_H) \leq 0$ for a large enough φ_H does follow directly. From the manufacturer's optimization it is clear that when φ_L is small enough, setting a w that induces $\bar{c} > 0$ is optimal and by (1.1) also feasible for the manufacturer. Hence, when φ_L is small enough $\bar{c}_L > 0$. □

On the supply side I have shown that when consumers follow the strategy of no search then only a price of 1 is charged in the high state and prices v and 1 are charged in the low state. Next, I must show that no searching is a best response for consumers. For notational clarity, define

$$w_L \equiv w(\varphi_L), \qquad w_H \equiv w(\varphi_H)$$
$$\bar{c}_L \equiv \bar{c}(w_L, \varphi_L), \quad \bar{c}_H \equiv \bar{c}(w_H, \varphi_H)$$

Low valuation consumers expect no prices strictly below v in either state and will either purchase on their first price draw or exit. High valuation consumers will assign a likelihood to the high state conditional on the price they see according to

$$\lambda(p) = \begin{cases} 0 & \text{if } p = v \\ \lambda_0 \, \frac{F(1-w_H)}{F(1-w_L)-F(\bar{c}_L)} & \text{if } p = 1 \\ \lambda_0 \, \frac{1-F(1-w_H)}{1-F(1-w_L)} & \text{if } p = \infty \\ \lambda_0 & \text{for all other } p \end{cases} \tag{1.14}$$

Likelihoods at equilibrium prices are computed as the product of the prior likelihood λ_0 and the ratio of the probabilities of seeing the price in either state. Prices off the equilibrium path are by assumption ignored by consumers when forming beliefs. Note that likelihood λ translates into belief $\mu = \frac{\lambda}{\lambda+1}$.

A high type consumer whose lowest observed price is p and who holds belief μ has a value function recursively defined by

$$V(p, \mu) = \max\Big\{0,\ 1-p,\ E[V(p', \mu')|p, \mu] - s\Big\} \tag{1.15}$$

Claim I.3. *There exists a high enough belief $\bar{\mu}$ so that whenever $\mu > \bar{\mu}$, the continuation value to searching $E[V(p', \mu')|p, \mu] - s < \max\{0, 1-p\}\ \forall\ p$.*

Proof of Claim I provide an upper bound for the continuation value to searching:

$$E[V(p', \mu')|p, \mu] - s \leq \mu(\max\{1-p, 0\}) + (1-\mu)(1-v) - s$$

If the state is high, the consumer will not see a price below 1 and the highest payoff she can obtain is to accept p if it is less than 1 else exit. If the state is low, the highest payoff the consumer can get is if she observes and accepts price v. For large enough μ, it must then be that

$E[V(p', \mu')|p, \mu] - s < \max\{0, 1 - p\}$. □

To restate the claim, for any price p once consumers are convinced enough the state is high they will not search and either purchase or exit.

Claim I.4. *There exists a* $\bar{\lambda}$ *so that for any* $\lambda_0 > \bar{\lambda}$, $\mu(p) > \bar{\mu}$ $\forall$ p.

Proof of Claim The proof follows from equation I have thus shown that for large enough λ_0 no price will induce search and this concludes the proof of the lemma. □

CHAPTER II

Optimal Timing of Selection Contests

2.1 Introduction

Going into the 2008 Beijing Olympics hopes were particularly high for American swimmers Michael Phelps and Katie Hoff, both world record holders and slated to compete in numerous events. Yet at the Olympics while Phelps met and exceeded expectations Hoff under-performed. Why the disparity in the two athletes' performances? Putting aside explanations relying on swimmer-specific idiosyncracies, their performances can be understood by considering the athletes' strategic environment. Swimmers can peak only for a short period and they time their peak for a particular date. Each could peak closer to the trials to improve their chances of making the team but, by doing so hurt their Olympic performance. Katie Hoff faced stiff domestic competition and was forced to peak closer to the trials; Michael Phelps maintained a comfortable lead over his American rivals and was able to peak closer to the Olympics.[1] For USA Swimming, Katie Hoff's mediocre performance was avoidable: had they chosen a significantly earlier date for the trials, she would have had the opportunity to recover and peak again at the Olympics.[2] But holding the trials early would have also come at a cost, since the best swimmer at the trials might no longer still be the best swimmer at the Olympics. In choosing the optimal time to hold the trials, USA Swimming had to balance the accuracy of their selection with the cost of not allowing swimmers sufficient time to recover.

This problem facing USA Swimming is one faced by many organizations: how to time the selection of an agent when waiting longer makes the selection more accurate but also more costly. Take as another example a political party choosing a candidate for a general election. Candidates

[1] In each event USA Swimming selects the top two performers. Going into the Olypmic trials, Michael Phelps' average lead over the eventual third place finishers was roughly twice that of Katie Hoff, in percentile terms.

[2] In fact Australia, another traditional swimming power, held their trials a full year in advance.

will attempt to position themselves on the political spectrum to match the views of their party's median voter, but in order to win the general election they would be best off at the median of the entire population. Given that shifting one's political platform takes time, holding a later primary hurts the chances of the eventual primary winner. At the same time, a lengthy primary election can reveal characteristics about a candidate which are independent of his or her platform but are important in determining electability in the general election.[3] In this sense a later primary ensures that the winner is the most likely of the group to win the general election. Party organizers must weigh the accuracy benefits of a later primary with the costs of a running a more extreme candidate in the general election.[4]

A similar tradeoff is also found in the workplace. A manager must select which of several analysts to promote to associate. Delaying the promotion decision allows the manager to ascertain which of his analysts is most highly skilled. However, the longer the promotion decision is delayed, the more time is spent by the eventual winner performing the tasks of analyst instead of the more productive tasks of an associate.

In all of these examples, a principal faces a group of agents whose types he does not know and must choose one to perform a task. To make this decision, the principal chooses a time at which all agents compete and uses the results of this competition to make the selection. Agents divert resources from their final task to improve their performance in the competition and the principal must keep this in mind when choosing the timing optimally. The aim of this paper is to explore this timing decision and to ascertain which factors induce the principal to choose a later more accurate contest and which favor an earlier and noisier selection.

We take two approaches to address this question. First we note that from the point of view of the principal, the choice of timing is indirectly a choice of an allocation rule and a corresponding incentive compatible transfer function. That is, for any given selection time, agents choose how many resources to divert to the contest from their final performance, i.e. their transfer, and in an equilibrium conditional on these actions agents will have some probability of winning, i.e. the allocation rule. The choice of timing can be thought of as a blunt tool in the more general prob-

[3]For example, if candidate is a particularly inspiring public speaker or is has damaging secrets from his or her past.

[4]The timing of the primary can be broadly understood as the time at which candidates begin to compete, for instance the first televised debate.

lem of choosing an optimal incentive-compatible screening mechanism, and our first approach is to solve that more general problem. Specifically, we consider a setting in which agents are privately informed about their types and simultaneously send costly signals to the principal, who in turn commits to an allocation rule as a function of the signals. The principal's payoff is the performance of his chosen agent, hence he wants to choose the best agent but suffers that agent's signaling cost. We show that ex-ante the principal prefers to allocate not necessarily to agents with the highest type but to those with the lowest hazard ratio, roughly speaking those agents facing the least competition. The optimal mechanism is stochastic, that is for some collections of signals the principal will allocate using a lottery between some or all of the agents. While this may seem suboptimal ex-post, a stochastic mechanism lessens the returns to costly signaling and reduces competition ex-ante. Thus, it is best to include some noise in the selection process.

The stochastic optimal allocation rule relies on commitment by the principal and may be difficult to implement in practice. Taken at face value, the mechanism requires the principal to sometimes not select the best agent despite having perfectly inferred all participants' types through their signals. The principal cannot always commit to this behavior, just as, for instance, a party chairman may not find it feasible to flip a coin between two candidates when one has received significantly more votes. On the other hand, the party chairman is able to commit to an early primary and this can indirectly infuse a stochastic component into the decision process. Given that a candidate's types evolves over time, the person chosen early is not necessarily the best one at the time of the general election.[5]

Restricting the principal to making only a timing decision then reduces the set of allocation rules that he can implement, and in our second approach we look for the the optimum in this reduced set. We consider a continuous time environment in which agents privately observe the evolution of their types over time and choose an effort level for the selection contest that comes at the expense of their final performance. We assume that the cost of effort decreases as the time between the selection and the final task grows so that agents are better able to recover given more time. In choosing a later selection time, the principal benefits from the option value of picking the agent with the best shocks but potentially pays the price in higher effort costs. We show that if agents' ability to divert resources from the final task is unrestricted, agents divert the same amount of

[5]Or more precisely, if an agent's type evolves so that it is best at the time of the final task, she may not have necessarily been chosen during the selection contest.

resources from the final task regardless of the timing of the contest. That is, as there is less time to recover between the selection and the final task and effort becomes costlier, all agents adjust their effort to exactly offset this. Hence, it is always optimal to select agents as late as possible.

This result sheds some light on the underlying mechanism and we treat it as a benchmark case. However, the assumption that agents are unrestricted in how many resources they can divert is strong and violated in many applications. For instance in swimming, even if athletes peak for the trials, they can once again peak for the Olympics given enough time to recover. By choosing a trials date sufficiently early the principal can thus restrict the amount of resources participants can divert. To capture this, we consider a simplified timing model in which the principal chooses between an early and a late contest. The early contest is sufficiently early so that agents will fully recover; hence, they are unable to divert resources from their final performance. The late contest is one in which an agent's ability to divert resources is unconstrained. We assume that all agents start with the same type at the early contest and since types evolve, each agent's type is expected to be distributed according to some function F at the time of the late contest. We show that the principal's payoff to holding a late contest is independent of the mean of F but grows in the dispersion of F. We also show that the effect of adding more agents to the late contest is ambiguous as there is a tradeoff between choosing a type from a higher order statistic but also inducing more competition. In addition, we consider selection contests of skill in which higher types have a natural advantage and prove this improves the returns to a later contest.

This paper is concerned with the question of how to time a selection contest, and answering this question overlaps with several literatures. A selection contest is in essence a screening mechanism, and as such our setting is strategically similar to those in the seminal works of Spence (1973) and Myerson (1981), as well as in the myriad papers spawned thereafter. Aside from establishing a general framework, Myerson (1981) also provides several techniques that we adopt to characterize equilibria. A point of departure from these classic works is that the designer can suffer the costs (or transfers) along with the agents. This idea is present in more recent work such as Hartline and Roughgarden (2008) and Chakravarty and Kaplan (2006). Both papers consider a setting in which agents' efforts are wasteful to the principal, with the latter explicitly characterizing the optimal mechanism for a benevolent social planner that wants to assign a good but suffers the sum of all the signaling costs of the agents. Our paper is motivated by a selfish designer who cares only about the costs expended by the agent that he chooses, yet we show that the two problems are strategically equivalent and the same mechanism maximizes both objectives. We re-interpret Chakravarty and Kaplan's result in the context of competition, showing that the optimal allocation rule assigns the good to those who have to compete the least. We also differ in spirit from Chakravarty and Kaplan (2006) in addressing the issues of commitment and implementation by focusing on the timing question in particular.

In the part of the paper that deals with the timing model, we ask questions that are often addressed in the field of tournament design, such as how to assign a winner (Lazear and Rosen (1981), Moldovanu and Sela (2001)) and how many participants to include (Fullerton and McAfee (1999)). The answers to these questions are quite different in our setting in which a principal wishes to minimize effort as compared to a tournament design setting in which the principal wants to maximize effort.

The rest of the paper proceeds with Section **??**, in which we present a model of costly screening. There we establish the equivalence to the general solution from Chakravarty and Kaplan (2006) and offer an interpretation for our particular application. In Section **??** we drop the assumption that a principal can commit to an allocation rule and explicitly model a principal that must choose a time for a selection contest. Then Section **??** concludes.

2.2 A Model of Costly Screening

We consider a setting in which a principal must choose one among N agents to perform a task. Each agent i has type θ_i which describes her ability to perform the task. Types are drawn privately and independently from distribution $F(\cdot)$ and are not observed by the principal. An agent i can signal her type by taking a publicly observable costly action $e_i \geq 0$. The cost of this action is that it diverts resources away from the agents' performance of the task conditional on being selected. For instance, e_i can represent how far a candidate in a primary election moves her platform away from that of the population's median voter; the cost of this move is a reduced chance of winning the general election and is only borne if the candidate wins the primary. An agent's performance conditional on being chosen is

$$y_i = \theta_i - e_i$$

Agents choose e_i simultaneously and the principal can commit to a mechanism

$$M : e \to x$$

in which the principal maps each action profile $e = \{e_1, ..., e_N\}$ into a profile of allocations $x = \{x_1, ..., x_N\}$, where x_i is the probability that agent i is selected. We refer to the vector $x^M(e)$ as the allocation rule under mechanism M. We explicitly allow for the allocation mechanism to be stochastic since a central question of our analysis is whether a stochastic mechanism is optimal. A mechanism M is feasible if for every profile e, $\sum_i x_i^M(e) \leq 1$.

Agent i's payoff is her performance on the task, and her VNM expected utility function is

$$u_i = x_i(\theta_i - e_i)$$

The principal's payoff is his chosen agent's performance, hence his VNM utility is

$$u_0 = \sum_i x_i(\theta_i - e_i) \ = \sum_i u_i$$

An equilibrium is characterized by two conditions: expected utility maximizing behavior of each agent conditional on mechanism M^*

$$e_i^*(\theta_i, M^*) = \arg\max_{e_i} E_{e_{-i}}[x_i^{M^*}(e_i, e_{-i})(\theta_i - e_i)] \tag{2.1}$$

and the principal's optimal choice of mechanism M^*

$$\begin{aligned} M^* &= \arg\max_M \ E_\theta \left[\sum_i x_i^M(e^*(\theta, M))(\theta_i - e_i^*(\theta_i, M)) \right] \\ &= \arg\max_M E_\theta \left[\sum_i u_i(\theta, M) \right] \end{aligned} \tag{2.2}$$

One way to think about the optimal allocation mechanism M^* is as a tradeoff between choosing accurately and inducing costly signaling. As extreme examples, an allocation rule $x_i(e) = \frac{1}{N}$ would choose randomly but induce $e_i = 0$ by all participants while $x_i(e) = \mathbb{I}\{e_i = \max\{e_1, ..., e_N\}\}$ would always select an agent with the highest action but would induce high signaling costs. As we will show, the optimal mechanism uses either the latter of these two allocation rules or is a lottery amongst those with the highest effort levels, depending on vector e.

We use a mechanism design approach to characterize the M^*, and we do so in two steps. First, we invoke the revelation principle to show that the allocation rule x^{M^*} is the same as an allocation rule x^{DM^*} which maximizes the sum of the expected utilities of the agents in a direct revelation mechanism. Then, we use a result in Chakravarty and Kaplan (2006) which characterizes x^{DM^*}.

From Myerson (1981), we know that if $(x^M(e), e^*(\theta, M))$ is an equilibrium in our setting then in a direct revelation game $(x^{DM}(\theta), \tau(\theta, x^{DM}))$ is a truth-telling equilibrium if $x^{DM}(\theta) = x^M(e^*(\theta, M))$ and $\tau(\theta, x^{DM}) = e^*(\theta, M)$.

$$\begin{aligned}
x^{M^*}(e) &= \arg\max_{x^M} \; E_\theta \left[\sum_i x_i^M(e_i^*(\theta_i, x^M))(\theta_i - e_i^*(\theta_i, x^M)) \right] \\
&= \arg\max_{x^{DM}} \; E_\theta \left[\sum_i x_i^{DM}(\theta)(\theta_i - \tau_i(\theta, x^{DM})) \right] \\
&= \arg\max_{x^{DM}} \; E_\theta \left[\sum_i u_i\left(\theta, x^{DM}\right) \right] \\
&= x^{DM^*}(\theta)
\end{aligned} \tag{2.3}$$

The revelation principle is invoked in going from the first to second line, equating choosing an allocation x^M conditional on agents playing equilibrium $e^*(\theta, x^M)$ to choosing allocation x^{DM} given that transfers $\tau(\theta, x^{DM})$ are constructed so that truth-telling is optimal.

Proposition II.1. *(From Chakravarty and Kaplan (2006), Proposition 1)*

Let $x^{DM^*}(\theta)$ be the allocation rule in a direct revelation mechanism that maximizes the sum of participants' expected utilities, as in (2.3). Let $\theta_M = \max\{\theta_1, ..., \theta_N\}$. There exists a family of disjoint intervals $I_k = [\theta_{k-}, \theta_{k+}]$, $k = 1, ..., K$ such that

$$\textit{if } \theta_M \in I_k \textit{ for some } k \textit{ then} \qquad x_i(\theta) = \begin{cases} 1/\#(\theta_j \in I_k) & \textit{if } \theta_i \in I_k \\ 0 & \textit{otherwise} \end{cases}$$

$$\textit{else if } \theta_M \notin I_k \textit{ for any } k \textit{ then} \qquad x_i(\theta) = \begin{cases} 1/\#(\theta_j = \theta_M) & \textit{if } \theta_i = \theta_M \\ 0 & \textit{otherwise} \end{cases}$$

We refer the interested reader to Chakravarty and Kaplan (2006) for a characterization of the intervals I_k along with a proof of the proposition. The authors call intervals I_k the "lottery" regions and the rest of the type space as the "contest" regions. Whenever the highest type is in a lottery region, the allocation is a random assignment among all others in the region. In what follows we provide some intuition for why such an allocation rule is optimal with an example.

Consider a setting with $N = 2$ agents and assume that the lowest type is $\underline{\theta} = 0$. Using the law of iterated expectations, we can rewrite the principal's objective in the direct mechanism from equation (2.3)

$$E_\theta\Big[\sum_i u_i(\theta, x)\Big] = 2\int_{\theta_i} E_{\theta_j}\Big[u_i(\theta_i, x_i)\Big] f(\theta_i) d\theta_i \tag{2.4}$$

In other words, the principal wants to maximize the ex-ante expected utility of a participant. Following Myerson (1981), an allocation rule $x(\theta)$ can be obtained in equilibrium if it satisfies the following four constraints:

$$E_{\theta_j}\Big[u_i(\theta_i, x_i)\Big] = E_{\theta_j}\Big[u_i(\underline{\theta}, x_i)\Big] + \int_{\underline{\theta}}^{\theta_i} X(\hat{\theta}) d\hat{\theta} \tag{2.5}$$

$$X(\theta_i) \quad \text{non-decreasing} \tag{2.6}$$

$$E_{\theta_j}\Big[u_i(\theta_i, x_i)\Big] \geq 0 \tag{2.7}$$

$$\sum_i x_i(\theta) \in [0, 1] \quad \text{and} \quad 0 \leq x_i \leq 1 \tag{2.8}$$

Conditions (2.5) and (2.6) are derived from first and second order conditions for truthful reporting, respectively. Condition (2.7) is a participation constraint and (2.8) is a feasibility constraint.

To solve the constrained optimization, we solve a relaxed program by ignoring (2.6), and then use a technique from Myerson (1981) called ironing to apply that constraint. Plugging (2.5) into the objective function (2.4), we obtain

$$\begin{aligned} & \max_x \ 2E_{\theta_j}\Big[u_i(\underline{\theta}, x_i)\Big] + 2\int_{\theta_i}\int_{\underline{\theta}}^{\theta_i} X(\hat{\theta}) f(\theta_i) d\theta_i \\ \Leftrightarrow \quad & \max_x \ 2\int_{\theta_i} X(\theta_i)(1 - F(\theta_i)) \ d\theta_i \\ \Leftrightarrow \quad & \max_x \ 2\int_{\theta_i}\left(\int_{\theta_j} x_i(\theta_i, \theta_j) f(\theta_j) \ d\theta_j\right)(1 - F(\theta_i)) \ d\theta_i \end{aligned} \tag{2.9}$$

The second line follows because, with non-negative transfers, type $\underline{\theta} = 0$ gets a utility of at most zero, and, by the participation constraint, must get exactly zero. Next we claim without proof that (2.8) binds so that $x_1(\theta_1, \theta_2) + x_2(\theta_1, \theta_2) = 1$ and that we can restrict attention to symmetric allocations, so $x_1(a, b) = x_2(b, a)$. This allows us to rewrite the objective as

$$\max_x \; 2 \int_{\theta_i} \int_{\theta_j \leq \theta_i} x_i(\theta_i, \theta_j) f(\theta_j)(1 - F(\theta_i)) + (1 - x_i(\theta_i, \theta_j)) f(\theta_i)(1 - F(\theta_j)) \; d\theta_j \; d\theta_i \tag{2.10}$$

Expression (2.10) can be maximized pointwise, and by inspection the optimal allocation rule must satisfy

$$x_1(\theta_1, \theta_2) = \begin{cases} 1 & \text{if} \quad \frac{1-F(\theta_1)}{f(\theta_1)} \geq \frac{1-F(\theta_2)}{f(\theta_2)} \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

We can think of θ_i as having a "virtual type" given by the inverse hazard ratio $\frac{1-F(\theta_i)}{f(\theta_i)}$. The hazard ratio is a measure of how competitive the neighborhood is around a particular type, and the principal wants to allocate probability to types that will not be forced to compete very hard.[6]

This allocation rule will generally not satisfy condition (2.6), as higher types will tend to have lower virtual types. For instance, if θ_i is distributed uniformly on $[0, 1]$ then (2.11) results in interim probabilities given by $X(\theta_i) = 1 - \theta_i$, which is everywhere decreasing. To account for this, one must use a procedure similar to the ironing algorithm outlined in Myerson (1981). The basic idea is that over regions in which (2.11) causes $X(\theta_i)$ to be decreasing, one must increase the allocation to higher types at the expense of lower types, and the optimal way to do so is to identify intervals such that whenever two types are from the same interval their allocation is decided by a random lottery.

2.3 Timing Models

A key assumption in Section **??** is that the principal commits to the allocation rule and in many applications this assumption is difficult to defend. For example, imagine a party chairman flipping a coin between two candidates despite one receiving more votes than the other, or holding the US Olympic trials and then rolling a die to determine which of the top six performers goes to the

[6]This is in contrast to a principal who maximizes revenue and defines virtual types by $\theta_i - \frac{1-F(\theta_i)}{f(\theta_i)}$. This principal prefers agents with high hazard ratios so that they compete away their information rents.

Olympics. Aside from having incentive to choose the revealed-best agent, there is also institutional pressure to do so.

Yet while the principal is pressured to choose the winner of the selection contest, he can make ex-ante decisions about the design of the contest so that the winner is not necessarily the highest type. We consider a setting in which the decision is about *when* to hold the selection contest. Agents have types that evolve stochastically over time, hence an earlier selection makes it less likely that the chosen agent will be best at the time of the task. However agents can also divert resources toward the selection contest, and the less time between selection and the task the costlier that diversion. The tradeoff in the timing decision is analogous to that in the previous section: accurate selection versus inducing harmful effort. In essence, the timing structure now restricts the set of available stochastic allocation rules and we look for the "second-best" optimum within this set.

2.3.1 A Model in Continuous Time

A principal commits to a time $t \in [0,1]$ in which to hold a selection contest. There are N agents and all have type θ_0 at time $t = 0$. Each agent's type evolves independently and stochastically over time such that for any $t > t'$, $\theta_i(t)$ is a mean preserving spread of $\theta_i(t')$. Let $F(\theta_i|t)$ denote the distribution of $\theta_i(t)$ and assume that for any t the support is always bounded weakly above zero and below infinity.[7] As an example, one may think of $\log(\theta_i(t))$ following Brownian motion, though many other stochastic processes would also be admissible.

Each agent privately observes the evolution of her type and at time t chooses action $e_i \in [0, \infty]$, which we will refer to as effort. The agent that wins the selection contest is the one with the highest score $s(e_i, \theta_i)$. For now, assume the score increases in effort only[8] so that

$$s(e_i, \theta_i) = s(e_i).$$

[7] This is an assumption made for mathematical convenience. In our setup, an agent with a negative type would choose not to participate and to compute equilibrium strategies would require accounting both for the distribution of types and the number of participants.

[8] An agent's score in the selection contest can depend both on her effort and her type. For example, in an Olympic trial a swimmer's chance of winning depends not only on how close she is to her peak but also on her ability. The assumption here is made to match the setup in the previous section; later we consider how relaxing this assumption affects the timing decision.

Each agent's action e_i comes at a cost $c(e_i, t)$ to the agent's final task performance

$$y_i = \theta_i(1) - c(e_i, t).$$

Note that an agent's final task performance depends on her type in period $t = 1$. However by assumption $E[\theta_i(1)|\theta_i(t)] = \theta_i(t)$, hence from the risk-neutral perspectives of both the agent and the principal, choosing type θ_i at time t has the same payoff as choosing type θ_i at time $t = 1$.

The cost function $c(e_i, t)$ is weakly increasing and continuous in effort and in time. Agents that choose $e_i = 0$ divert no resources from the final task, thus $c(0, t) = 0$ for any t. The set of feasible costs at any time is $[0, \bar{c}(t)]$, where $\bar{c}(t) \equiv \lim_{e\to\infty} c(e, t)$. Note that since c_e and c_t are both weakly increasing, it must be that $\bar{c}(t)$ is weakly increasing as well. As before, agent i's expected utility is $u_i = x_i y_i$ and the principal's expected utility is $u_0 = \sum_i x_i y_i$.

For a given selection time t, agents solve

$$\begin{aligned} & \max_{e_i} \ \Pr\Big(s_i(e_i) > \max_{j\neq i} s_j\Big)\Big(\theta_i - c(e_i, t)\Big) \\ \Leftrightarrow \quad & \max_{c_i \in [0,\bar{c}(t)]} \ \Pr\Big(s_i(e_i(c_i, t)) > \max_{j\neq i} s_j\Big)\Big(\theta_i - c_i\Big) \end{aligned} \tag{2.12}$$

The selection contest can then be interpreted as a first price auction in which each agent chooses bid function $c(\theta_i, t)$ bound by the budget constraint $\bar{c}(t)$. The principal anticipates expected equilibrium behavior for each time t and solves

$$\max_t \ E_{F(\theta_i,t)} \sum_i \Big(\theta_i(t) - c(\theta_i, t)\Big)$$

The timing choice is then a choice of a distribution of bidders $F(\theta_i, t)$ along with their corresponding bidding strategies $c(\theta_i, t)$. A later selection contest has the benefit of sampling from a distribution with higher variance but can also be associated with higher effort costs. We now examine this tradeoff in more detail.

Contests in Which $\bar{c}(t)$ Does Not Bind

First as a benchmark we examine the optimal timing decision when at every t, $\bar{c}(t)$ is large enough so that it does not bind.[9] In this situation we will show that in the timing decision there is no tradeoff for the principal. In holding a later contest the effort cost of the winner remains constant while the benefit of the option value of waiting continues to accrue. Thus, waiting until the end is optimal.

Proposition II.2. *Suppose $\bar{c}(t)$ is sufficiently large at every t so as not to bind. Then the optimal selection time is $t = 1$.*

Proof Sketch

Given that $\bar{c}(t)$ does not bind it must be that for any realization of types $(\theta_1(t), ..., \theta_N(t))$ the highest type wins for sure. This implies that type $\theta_i(t)$ has an interim probability of winning $X(\theta_i(t)) = F^{N-1}(\theta_i, t)$. Using the analog of expression (2.9) for N agents, we can write the principal's payoff as

$$u_0(t) = N \int_{\theta_i} F^{N-1}(\theta_i, t)(1 - F(\theta_i, t)) \, d\theta_i$$

We need to show that $\frac{\partial u_0}{\partial t} > 0$ by making use of the fact that as t increases the distribution F undergoes a mean-preserving spread. Roughly speaking a mean preserving spread makes bidding less competitive by making types more disperse and increases the expected type of the winner. Thus, the expected surplus of the winner also increases. See Appendix A for details. $\square$

Contests in Which $\bar{c}(t)$ May Bind

From Proposition II.2 we learn that when the cost of effort in the selection contest is unconstrained agents will exactly offset the cost savings of an earlier selection contest with higher effort, leaving the principal unable to affect the effort costs by his choice of timing. However, in many applications when the principal chooses the selection contest sufficiently early, even in maximizing their selection contest performance agents are only able to incur a limited cost to their final task performance. For instance, imagine an Olympic trials that takes place a full year prior to the Olympics. An athlete can peak at the trials and, with a full year to recover, peak again at the Olympics. Similarly, a political candidate can move her platform all the way to that of the party's median voter and still have sufficient time until the general election to move it back to the population median. In both

[9] For instance $\bar{c}(t) \geq \max_\theta supp(F(\theta, t))$

situations, by running the selection contest early enough the principal restricts the agents' ability to incur costs.

We accommodate for this by now considering a setting in which in some periods the cost constraint binds. In such a period t the equilibrium in the selection contest is characterized by a cutoff type $\hat{\theta}_i$ such that $c(\theta_i, t)$ is the solution to (2.12) for $\theta_i \leq \hat{\theta}_i$ and $c(\theta_i, t) = \bar{c}(t)$ for $\theta_i > \hat{\theta}_i$. The cutoff type is indifferent between the two bids. If the cutoff type bids according to (2.12), she wins if and only if all other agents have types below $\hat{\theta}_i$. Hence this cutoff type's interim probability of winning is $X_i(\hat{\theta}_i, t) = F^{N-1}(\hat{\theta}_i, t)$. If she were to bid $\bar{c}(t)$ she would win against all types below $\hat{\theta}_i$ and tie with any types above $\hat{\theta}_i$. With N agents there are several combinations in which ties can occur and the interim probability of winning is given by

$$\bar{X} \equiv \sum_{k=0}^{N-1} \frac{1}{k+1} \binom{N}{k} F(\hat{\theta}_i)^{N-1-k} (1 - F(\hat{\theta}_i))^k \tag{2.13}$$

Note that $\bar{X}(\hat{\theta}_i) > F^{N-1}(\hat{\theta}_i, t)$ since the possibility of tying with higher types makes the interim probability of winning jump discretely. When $N = 2$ the expression simplifies to

$$\bar{X}(\hat{\theta}_i) = F(\hat{\theta}_i) + \frac{1}{2}\Big(1 - F(\hat{\theta}_i)\Big)$$

The cutoff type $\hat{\theta}_i$ is then defined implicitly by the indifference condition

$$F^{N-1}(\hat{\theta}_i, t)(\hat{\theta}_i - c(\hat{\theta}_i, t)) = \bar{X}(\hat{\theta}_i)(\hat{\theta}_i - \bar{c}(t)) \tag{2.14}$$

In order for the equality to hold, it must be that when bidding the ceiling, the discrete jump in the interim probability of winning is accompanied by a corresponding discrete jump in the bid; thus $\bar{c}(t) > c(\hat{\theta}_i, t)$. Figure 1 depicts an example of a selection contest equilibrium with a binding ceiling constraint.

Having characterized what happens when the bid ceiling binds in the selection contest, we will argue shortly that without additional structure the timing of the selection contest is a complex problem that does not easily lend itself to comparative statics. To this end, we first ask a simpler question: Does a lower bid ceiling make the principal better off? That is if it were the case that an earlier selection contest reduces the maximal cost agents can incur and if we ignore the lost option

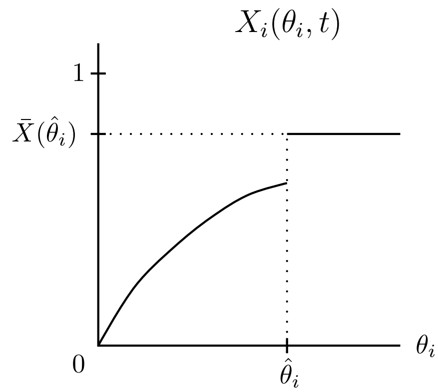

Figure 2.1: Interim allocations with a binding ceiling

value of waiting, does the principal benefit from an earlier contest?

The answer is not necessarily, and the reasoning behind this answer comes from the discussion of the optimal mechanism in Section 2.2. Recall that equation (2.10) allows us to express the payoff to the principal in any incentive compatible mechanism only in terms of the allocation rule. Recall also that ceteris paribus, in a collection of agents $(\theta_1, ..., \theta_N)$ the principal wants to assign all the probability to the agent with the highest virtual type $\frac{1-F(\theta_i)}{f(\theta_i)}$. In this sense, a lower bid ceiling is just a different allocation rule, and depending on the distribution of types F, not necessarily a better one.

Lemma II.1. *Depending on the distribution of types F, the principal can be either better or worse off with a lower bid ceiling. That is $\frac{\partial u_0}{\partial \bar{c}}$ is unsigned.*

See Appendix B for a proof of Lemma II.1.

Lemma II.1 highlights that the optimal timing problem is difficult to solve in general. Waiting to hold the contest always involves the benefit of cherry-picking those agents who have received a positive shock. However, the expected cost incurred by the winner in the selection contest can either increase or fall with time and, moreover, this effect can be non-monotonic. We now impose some additional structure on the problem that will provide tractability while still preserving the central tradeoffs.

2.3.2 A Two-Period Model

The principal now has only two choices of timing, an early contest at $t = 0$ or a late contest at $t = 1$. The early contest is defined as early enough so that agents can fully recover and peak again for the final task, hence $\bar{c}(0) = 0$. As before at $t = 0$ all agents have type θ_0. Since all agents have the same type and will expend no costs, it follows that the payoff to the principal from holding the selection contest in period $t = 0$ is

$$u_0(t = 0) = \theta_0$$

In the late contest agents are distributed according to $F(\theta_i, 1) = F(\theta_i)$. Recall that since types evolved over time through a mean-preserving stochastic process, $E[\theta_i] = \theta_0$. We assume that $\bar{c}(1)$ is large enough so as not to bind. This principal's payoff from holding the selection contest in period $t = 1$ is

$$u_0(t = 1) = N \int_{\theta_i} F^{N-1}(\theta_i)(1 - F(\theta_i))\ d\theta_i \tag{2.15}$$

Note that $u_0(t = 0)$ is a fixed value and thus in examining the timing decision we can solely focus on what affects $u_0(t = 1)$. For example, what about the shape of F makes a later selection more appealing? What happens to the benefit of a later selection when there are more participants? We provide first an example with a parameterized uniform distribution and then a series of lemmas to answer these questions more generally.

As a benchmark, first consider a distribution $F(\theta_i)$ that is uniform on $[\theta_0 - \alpha,\ \theta_0 + \alpha] \subset \mathbb{R}^+$. The expected surplus to the principal is

$$\begin{aligned} u_0(t = 1) &= N \int_{\theta_0-\alpha}^{\theta_0+\alpha} \left(\frac{\theta_i - (-\alpha + \theta_0)}{2\alpha}\right)^{N-1} \frac{\alpha + \theta_0 - \theta_i}{2\alpha}\ d\theta_i \\ &= \frac{2\alpha}{N+1} \end{aligned}$$

This surprisingly simple expression yields immediate comparative statics. The benefit to a later selection contest is (i) independent of θ_0, (ii) increases in the amount of noise, and (iii) decreases in the number of participants. Intuitively, increases in θ_0 get competed away by all the agents; a greater α adds to the option value of waiting; and a larger number of participants eats away at the winner's surplus through increased competition. In what follows, we will show that (i) and

(ii) hold in general, while there is a countervailing effect that makes (iii) depend on the distribution.

First, we show that the payoff to the later contest depends only on the shape of F and is not a function of where F is centered.

Lemma II.2. *For some $a > 0$, let $G(\theta_i) = F(\theta_i - a)$ so that distribution G is obtained by shifting F to the right by a. Then $u_0(t = 1|F) = u_0(t = 1|G)$.*

Proof. Follows from applying a change of variables to (2.15). □

One can think of an agent's type as composed of a common component θ_0 and an idiosyncratic component $\theta_i - \theta_0$. The lemma states that a later selection contest erodes away the common component through competition, and while this relies on the fact $\bar{c}$ does not bind, the qualitative message still holds when the assumption is relaxed. One implication is that if a principal can make a costly investment which can improve all agents' types, he should weigh this investment carefully since the eventual winner will have lost much of its benefit through competition. On the other hand, when holding the selection contest early the principal gets the full return on such an investment.

While the mean of F has no effect on the payoff of a later contest, given that there is an option value associated with waiting, we may expect that when F is more disperse a later contest is more beneficial. However, the shape of the distribution also effects the level of competition, thus the result is not immediate and requires a more careful analysis.

Lemma II.3. *If $G(\theta_i)$ is a mean preserving spread of $F(\theta_i)$ then $u_0(t = 1|G) \geq u_0(t = 1|F)$.*

Lemma II.5 is proven in Appendix A.

This result is a generalization of the finding in the uniform case in which $u_0(t = 1)$ is increasing in α. Note that MPS is a partial ordering over distributions and that the converse of the lemma is not true. That is, one can construct examples in which $u_0(t = 1|G) > u_0(t = 1, F)$ but G is not a mean preserving spread of F. The two lemmas combine into one of our main propositions.

Definition II.1. *The distribution of Y is a spread of the distribution of X iff there exists a random variable Z and constant c such that (i) $E[Z|X] = c$ for all values of X and (ii) $Y \overset{d}{=} X + Z$.*

Proposition II.3. *If $G(\theta_i)$ is a spread of $F(\theta_i)$ then $u_0(t = 1|G) \geq u_0(t = 1|F)$.*

Next we consider how the number of agents N figures into the timing decision. The payoff to the early $t = 0$ contest is independent of N. However, at $t = 1$ the number of agents matters in two ways. On one hand, the identity of the highest of N draws is on average higher when N is higher. On the other hand, an increase in the number of agents makes the contest more competitive, inducing the eventual winner to bid a higher cost. As we will show, either effect can dominate.

We consider a parameterized family of distributions $F(\theta_i) = \theta_i^a$ with support on $[0, 1]$ and $a > 0$. Using equation (2.15) we can compute the principal's payoff in the later contest as a function of N.

$$\begin{aligned} u_0(N) &= \int_0^1 \theta_i^{a(N-1)}(1-\theta_i^a)\, d\theta_i = \frac{aN}{(1+a(N-1))(1+aN)} \\ u_0(N+1) - u_0(N) &= \frac{a}{(1+a(N-1))(1+a(N+1))(1+aN)}(1-a(N+1)) \\ &\geq 0 \ \text{ if and only if } a \leq \frac{1}{N+1} \end{aligned}$$

When a is sufficiently small the principal is better off with an extra agent, otherwise he is better off with one fewer agent. With more agents the field becomes more competitive but also the odds of drawing a very good agent increase, and the margins of both of these effects are sensitive to the shape of the distribution. In the particular family that we have chosen, for small values of a most agents are already crowded in a small interval at the low end of the distribution and the marginal increase in competition from an additional agent is negligible relative to the marginal increase in the likelihood that an extra agent will draw a much higher type.

Lemma II.4. *The principal's payoff to the later contest can increase or decrease with the number of participants.*

2.3.3 Contests of Skill

Thus far we have assumed that the selection contest is one of will in the sense that higher types win because they are willing to pay a higher cost in order to win. But often, selection contests are also contests of skill so that of two agents putting in the same effort, the agent with the higher type will score higher in the contest. Thus, in this section we assume that $s(e_i, \theta_i)$ is increasing in the second argument and analyze how this change affects the optimal timing decision.

First, note that although it may seem intuitive that reducing agents' cost of obtaining a high score in the selection contest leads to lower costs and better final task performance, we have already implicitly shown that generally this is not the case. Specifically, note that in Section **??** no assumptions were placed on the function $s(e_i)$ other than it being increasing. Hence, if we were to change the contest so that we have a new scoring function $\tilde{s}(e_i) = s(2e_i)$, thus making a given score half as expensive, the equilibrium utility to the principal would remain unchanged. We will show that it is not the absolute value of $s(e_i, \theta_i)$ that matters but rather how it varies with θ_i.

For clarity, we continue to evaluate the two period case, specifically the contest at $t = 1$ with a $\bar{c}$ that does not bind. Assume that $s_2(e_i, \theta_i) > 0$. Let $c(s, \theta_i)$ be the cost borne by type θ_i in obtaining a score of s in the selection contest. An agent's expected utility is

$$u_i(s_i, \theta_i) = \Pr\left(s_i > s_{j \neq i}\right)\left(\theta_i - c(s_i, \theta_i)\right)$$

Claim II.1. *There is a separating equilibrium in which $s(\theta_i)$ is strictly increasing, thus the interim probability of winning is $X_i(\theta_i) = F^{N-1}(\theta_i)$.*

Proof Sketch

In the original setting with $s(e_i, \theta_i) = s(e_i)$, this result was guaranteed by the single crossing property since higher types have a higher marginal willingness to pay and the same marginal cost. In our setting, higher types still have a higher marginal willingness to pay and a smaller marginal cost. □

Since the equilibrium is separating, we can use the standard direct mechanism tools to compute agents' expected equilibrium utility. Given some equilibrium function $s(\theta_i)$, let $c(\tilde{\theta}, \theta_i) \equiv c(s(\tilde{\theta}), \theta_i)$ be the cost to type θ_i of choosing the equilibrium score of type $\tilde{\theta}$. An agent's expected utility can be rewritten in terms of her type and her report is

$$\begin{aligned} u_i(\tilde{\theta}, \theta_i) &= F^{N-1}(\tilde{\theta})(\theta_i - c(\tilde{\theta}, \theta_i)) \\ \frac{du_i}{d\theta} &= \frac{\partial u_i}{\partial \tilde{\theta}}(\theta_i, \theta_i) + \frac{\partial u_i}{\partial \theta_i} \\ &= 0 + F^{N-1}(\theta_i) - c_2(\theta_i, \theta_i) \end{aligned} \tag{2.16}$$

That $\frac{\partial u_i}{\partial \tilde{\theta}}(\theta_i, \theta_i) = 0$ follows because reporting one's own type is a best response. Expression (2.16) is similar to the incentive constraint from Myerson (1981), with the last term accounting for the heterogeneous costs. An agent's expected utility in the contest is then

$$u_i(\theta_i) = u_i(\underline{\theta}) + \int_{\hat{\theta}} F^{N-1}(\hat{\theta}) - c_2(\hat{\theta}, \hat{\theta})\ d\hat{\theta}$$
$$= \int_{\hat{\theta}} F^{N-1}(\hat{\theta}) - c_2(\hat{\theta}, \hat{\theta})\ d\hat{\theta} \tag{2.17}$$

The last line follows because type $\underline{\theta}$ wins with probability zero. Since $s_2(e_i, \theta_i) > 0$, it must be that $c_2(s_i, \theta_i) < 0$. Comparing (2.5) with (2.17), it is evident that when higher types are more skilled in the selection contest, the interim expected utility of all agents and thus the expected utility of the principal are higher.

Proposition II.4. *The more is the selection contest one of skill, the more beneficial it is to hold it at $t = 1$. That is, given two selection contests $s^A(e_i, \theta_i)$ and $s^B(e_i, \theta_i)$, if $s_2^A > s_2^B$ at every point, $u_0(t = 1|s_A) > u_0(t = 1|s_B)$.*

2.4 Conclusion

We consider the decision of a principal that times a selection contest for a group of agents whose types he does not know. The principal uses the outcome of the contest to choose the agent best suited to perform a task, and any effort that agents put forth in order to win the contest hurts their task performance. Agents' types evolve over time making a later selection more accurate, however as time passes effort in the contest becomes costlier to the agents' task performance conditional on being selected. We examine the determinants of the principal's decision in light of this inherent tradeoff.

Considering the timing problem directly, we find that while effort is less costly in earlier contests much of these savings are wiped out by agents simply putting in more effort. Hence it is not enough for a principal to pick a slightly earlier contest; he must pick one early enough so as to physically constrain agents from being able to divert resources from the final task. We show that the benefits of a later contest are only a function of the dispersion of agents' types and not a function of the average of their types, which means a later contest is only optimal if the dispersion of ability is sizeable relative to the average ability. We also show that including more agents in a later contest is not necessarily beneficial, as doing so trades off drawing a potentially higher type against the effects of increased competition. Lastly we show that in selection contests in which skill gives agents an edge, a later contest is more favored because the eventual winner is likely to incur a smaller cost. Improving all agents' ability in the selection contest however provides no benefit.

Recasting this problem as one of designing an optimal screening mechanism provides several insights. Namely, due to the fact that competition is costly for the principal, he wants to design a noisy contest so that the incentive for effort is muted. Along the same lines, when taking the effects of competition into consideration, the principal wants to design a mechanism that favors those agents that are "isolated" in the distribution and thus compete away relative little of their surplus.

The optimal screening mechanism provides a first best benchmark but is not always implementable, as it requires a principal to commit to actions ex-ante that he may not want to take ex-post. The timing games on the other hand require no commitment by the principal, as every action he takes is best conditional on his beliefs. Yet timing games are just a subset of the bigger set of implementable mechanisms, and an important extension would be to characterize this set and the optimum within it.

2.5 Appendix

Appendix A

Proposition II.2 and Lemma II.5 follow directly from this lemma:

Lemma II.5. *If $G(\theta_i)$ is a mean preserving spread of $F(\theta_i)$ then*

$$\int_{\theta_i} G(\theta_i)^{N-1}(1 - G(\theta_i))d\theta_i \geq \int_{\theta_i} F(\theta_i)^{N-1}(1 - F(\theta_i))d\theta_i$$

Proof. Rothschild and Stiglitz (1970) have shown that if G is a mean preserving spread of F then there exists a sequence F_i so that $F_0 = F$ and $F_i \to G$ such that F_n and F_{n+1} are separated by what the authors call a "single Mean Preserving Spread" (single MPS). Our goal is then simplified to prove the proposition only for F and G separated by a single MPS.

We restate the Rothschild and Stiglitz (1970) definitions here. It is useful to define the function

$$s(x) = \begin{cases} \alpha & \text{for } a < x < a+t \\ -\alpha & \text{for } a+d < x < a+d+t \\ -\beta & \text{for } b < x < b+t \\ \beta & \text{for } b+e < x < b+e+t \\ 0 & \text{otherwise} \end{cases}$$

with

$$\alpha, \beta \geq 0$$
$$a \leq a+t \leq a+d \leq a+d+t \leq b \leq b+t \leq b+e \leq b+e+t$$
$$\alpha d = \beta e$$

An example of such a function is depicted in Figure 2.2. When the function $s(x)$ is added to a probability density function, $f(x)$, it shifts α probability mass down a distance d and β probability mass up a distance e, and since $\alpha d = \beta e$, this preserves the mean.

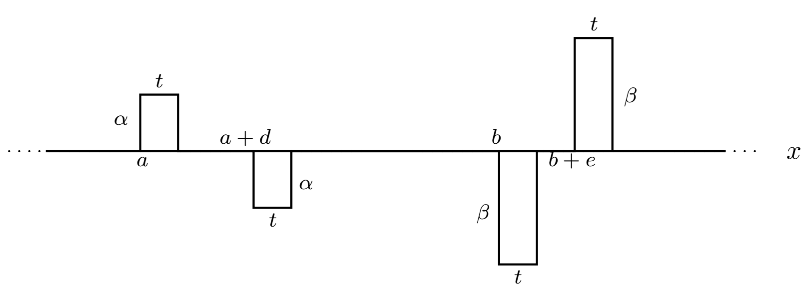

Figure 2.2: A single MPS $s(x)$

A couple of things to note here. Given that probability density functions are defined so $g(x) = f(x)+s(x)$, it must be that $G(x) > F(x)$ for $a < x < a+d+t$ and $G(x) < F(x)$ for $b < x < b+e+t$; for all other x, it must be that $G(x) = F(x)$. Furthermore,

$$\int_a^{a+d+t} G(x) - F(x)\ dx = \int_b^{b+e+t} F(x) - G(x)\ dx \equiv \Delta$$

Thus to go from F to G we first add a mass of Δ over the interval $[a, a+d+t]$ and then take that mass away over the interval $[b, b+e+t]$.

Claim II.2.

$$\begin{aligned}&\int_a^{a+d+t} G(x)^{N-1}(1-G(x)) - F(x)^{N-1}(1-F(x))\ dx \\ > \quad & (F(a+d+t)+\Delta)^{N-1}(1-(F(a+d+t)+\Delta)) - F(a+d+t)^{N-1}(1-F(a+d+t)).\end{aligned}$$

and

$$\begin{aligned}&\int_b^{b+e+t} G(x)^{N-1}(1-G(x)) - F(x)^{N-1}(1-F(x))\ dx \\ < \quad & (F(b)+\Delta)^{N-1}(1-(F(b)+\Delta)) - F(b)^{N-1}(1-F(b)).\end{aligned}$$

Proof. Follows from the fact that the function $F^{N-1}(1-F)$ is concave in F and maximized at $F = \frac{1}{2}$. □

Intuitively consider maximizing the function $\int F(\theta)^{N-1}(1-F(\theta))d\theta$ by adding a mass to a particular realization of $F(\theta)$. Because $F^{N-1}(1-F)$ is concave in F and maximized at $F = \frac{1}{2}$ the best place to add this mass is as far left as possible. Conversely, if we were to be removing a mass, the best place to do so would be as far right as possible. □

Putting the inequalities from the claim together, an immediate corollary is that

$$\int_a^{a+d+t} G(x)^{N-1}(1-G(x)) - F(x)^{N-1}(1-F(x))\ dx \geq \int_b^{b+e+t} G(x)^{N-1}(1-G(x)) - F(x)^{N-1}(1-F(x))\ dx$$

which implies that $\int G(x)^{N-1}(1-G(x))\ dx > \int F(x)^{N-1}(1-F(x))\ dx$. This concludes the proof of the lemma. □

Appendix B

Lemma II.1 is proven below.

Proof. The principal's expected payoff in this selection contest is given by equation (2.10)

$$\begin{aligned}u_0 &= N\int_{\underline{\theta}_i}^{\bar{\theta}_i} X(\theta_i, t)(1-F(\theta_i, t))d\theta_i \\ &= N\int_{\underline{\theta}_i}^{\bar{\theta}_i} F^{N-1}(\theta_i, t)(1-F(\theta_i, t))d\theta_i + N\int_{\hat{\theta}_i}^{\bar{\theta}_i} \left(\bar{X}(\hat{\theta}_i) - F^{N-1}(\theta_i)\right)(1-F^{N-1}(\theta_i))d\theta_i\end{aligned}$$

It follows from equation (2.14) that $\frac{\partial \hat{\theta}_i}{\partial \bar{c}} > 0$, hence it suffices to examine $\frac{\partial u_0}{\partial \hat{\theta}_i}$. Note that only the second term is a function of $\hat{\theta}_i$.

$$\frac{\partial u_0}{\partial \hat{\theta}_i} = N \left[\frac{\partial \bar{X}}{\partial \hat{\theta}_i} \int_{\underline{\theta}_i}^{\bar{\theta}_i} \left(1 - F^{N-1}(\theta_i)\right) d\theta_i \;-\; \left(\bar{X}(\hat{\theta}_i) - F^{N-1}(\hat{\theta}_i)\right) \left(1 - F^{N-1}(\hat{\theta}_i)\right) \right]$$

From equation (2.13),

$$\begin{aligned} \frac{\partial \bar{X}}{\partial \hat{\theta}_i} &= f(\hat{\theta}_i) \sum_{k=0}^{N-1} \frac{1}{k+1} \binom{N-1}{k} \left((N-1-k)F^{N-2-k}(1-F)^k - kF^{N-1-k}(1-F)^{k-1}\right) \\ &= f(\hat{\theta}_i) \sum_{k=0}^{N-1} \left(\frac{(N-1)!}{(N-2-k)!(k+1)!} \right) \left(\frac{1}{k+2} \right) F^{N-2-k}(1-F)^k \;>\; 0 \end{aligned}$$

The second line follows from the first through a grouping of terms. We see that $\frac{\partial \bar{X}}{\partial \hat{\theta}_i}$ is strictly positive and increases linearly in $f(\hat{\theta}_i)$. Fixing the function $F(\theta_i)$, the derivative $\frac{\partial u_0}{\partial \hat{\theta}_i}$ is positive for large values of $f(\hat{\theta}_i)$ and negative for small values. This is consistent with the idea that the principal wants to allocate to agents with high virtual types. An increase in $\hat{\theta}_i$ means that this type will now lose to people above her instead of tying with them. The larger is $f(\hat{\theta}_i)$, the lower is that agent's virtual type and the better off is the principal in allocating away from her. $\square$

CHAPTER III

A Model of Rational Speculative Trade

3.1 Introduction

A major discovery of economic theory is that two traders with a common prior and private information about a common value object cannot agree on a price at which to trade. This fact was initially conceived by Aumann (1976) and sparked a robust literature which continued to find this result in increasingly complex settings (seminal works include Milgrom and Stokey (1982), Kreps (1977), and Tirole (1982)). As it turned out, regardless of how one models a market rational agents simply cannot trade for purely speculative purposes. Tirole states that even in a dynamic setting, "speculation relies on inconsistent plans and is ruled out by rational expectations".

The robustness of these "no trade" theorems is quite surprising because in practice speculative trade is so common. Consider as a salient recent example the market for credit default swaps. In 2008, the notional amount of all the credit default swaps outstanding was $62 trillion dollars while the market for the underlying assets for which these contracts were written was roughly one twelfth the size. For every dollar owed by any bond issuer, there were twelve dollars of insurance purchased for it, hence in the vast majority of trades neither party had exposure to the underlying asset. In this regard credit default swaps are not unique; many other financial instruments are traded in volumes that seem significantly beyond what normal gains from trade would produce (see Odean (1999)).

Starting with Kyle (1985), a myriad of models has emerged offering a variety of behavioral explanations for the existence of what seems to be speculative trade. In this paper I present yet another explanation, however it is the only one in which agents are fully rational, share a common prior, and still willingly engage in speculate zero sum trade.

In the model agents all have private types, which in the context of a financial market can be thought of as an idea for a trading strategy. Agents can enter a market in which they are randomly matched with other agents and trade. The payoffs from two agents trading are such that a higher type extracts money from a lower type. The interaction can potentially have no gains from trade and be purely a zero sum game. A key assumption is that at the start agents do not know their own types. That is, an agent may have an idea but whether this idea works can only be ascertained through actually testing it out. As agents learn their types through trading, they may eventually discover that their idea is not good and exit the market. On the other hand, those who discover that their idea is profitable remain in the market and continue to make surplus.

In entering a trade an agent has two potential sources of value: the direct payoff and the option value of information. The presence of the informational benefit can induce an agent to enter a trade against an opponent who is on average of higher type. This is crucial, since in a steady state a potential entrant faces a market that is adversely selected against her.

The paper poses two questions. One is whether a steady state with entry exists in the extreme case of zero sum games. That is, can the informational benefit be strong enough to support a purely speculative market? To address this, I set up a model with discrete types in which learning is endogenous in the sense that the expected informational content of trades depends on the equilibrium distribution of the market. I demonstrate that for any parameter values a steady state with entry can be supported without any gains from trade.

The second question is that of volume. If some trades can occur in the purely speculative setting, what happens to the trade volume as some surplus is introduced? A discrete environment is not well suited for this inquiry, hence I amend the model so that agent's types are on a continuum. For tractability I simplify the learning rule so that information flows to agents at an exogenous rate. In the continuous setting I show it is still the case that a steady state with entry can always be supported when there is pure speculation. Then, I show that trading volume grows and is convex in gains from trade, a phenomenon I refer to as the speculative multiplier. The addition of surplus induces slightly lower types to remain in the pool and their presence becomes self-reinforcing.

The paper proceeds in four sections. Section 3.2 introduces the discrete model with endogenous learning, Section 3.3 amends the model to a continuum of types and exogenous learning, and Section 3.4 concludes.

3.2 Model with Discrete Types

Time passes discretely in periods $(..., t-1, t, t+1, ...)$. Each period a measure H of high type agents and L of low type agents are born and have the option to enter the existing pool. Those that choose not to enter remain out of the pool and receive a payoff of zero from that point on. Agents that enter are randomly matched with another participant in the pool and their interaction results in payoffs $u_i(\theta_i, \theta_j)$, in which θ_i and θ_j are the agents' types. An agent's lifetime utility is $U_i = \sum_t u_i(t)$. Each agent observes her payoff immediately after being matched and decides whether to remain in the pool for another round or exit. If she chooses to remain, the agent lives until the next round with probability $1-\delta$ and exits exogenously with probability δ.

Each agent is initially uninformed about her type and holds the common prior that she is a high type with probability $\frac{H}{H+L}$. If an agent enters the pool, she can learn about her type indirectly by observing the outcomes of her interactions and using Bayes rule to update beliefs.

I employ a steady state analysis, in which the makeup of the pool remains constant from one period to the next, both in the measure and distribution of agents. Let H_s and L_s denote the steady state measures of high and low types, respectively. In steady state the agent is uncertain only about her own type and perfectly anticipates the distribution of types in the pool.

The payoffs from the meeting of two agents are as follows:

$$u_a(a,b) = \begin{cases} 0 & a = b \\ x & a > b \\ -x & a < b \end{cases} \tag{3.1}$$

When two agents are of the same type, the payoff to both is zero; when one agent has a higher type then she extracts $x > 0$ from the agent with the lower type. Every meeting between agents is zero-sum and consistent with purely speculative trade.

The requirement that the outcome of any meeting is a function only of agents' types distinguishes this model from a typical model of dynamic matching and bargaining, in which agents's equilibrium payoffs depend also on the distribution of types in the pool. For example, in a bargaining model in which a buyer has unit demand her willingness to pay in any round is a function of her expected future offers, hence the distribution of agents affects her bargaining outcome. In contrast, agents in the model presented here should be thought of as risk-neutral speculative traders whose decision on whether to purchase an asset for a particular price is independent of previous or future purchases.

3.2.1 Steady State Equilibrium

A central question of this paper is whether a steady state can be supported in which agents make the initial decision to enter. It is not immediately clear that such an equilibrium can be sustained given the fact that the steady state pool is selected adversely against potential entrants. This is because only those agents that are sufficiently optimistic about their type remain in the pool. Still, I will find that the informational benefits of entry negate the initial expected trading losses.

It is useful to track agents' likelihood ratios rather than beliefs. With payoffs given by (3.1), the likelihood ratio of being a high type from one observation of payoff u is

$$\lambda(u) = \begin{cases} 0 & u = -x \\ \frac{H_s}{L_s} & u = 0 \\ \infty & u = x \end{cases} \tag{3.2}$$

and the likelihood ratio of a history of trade outcomes $(u_1, ..., u_T)$ is

$$\Lambda(T) = \lambda_0 \prod_{t=1}^{T} \lambda(u_t),$$

in which $\lambda_0 = \frac{H}{L}$. With any outright win $(u = x)$ or outright loss $(u = -x)$ an agent learns her type with certainty. In the case of a tie $(u = 0)$, the agent makes a statistical inference and the likelihood $\lambda(u = 0) = \frac{H_s}{L_s}$ plays an important role in the formulation of the equilibrium. If $H_s > L_s$ then a tie is good news in that an agent will be more convinced that she is a high type; conversely if $H_s < L_s$ then a tie is bad news. If it is the case that agents are exactly indifferent to entering for the first time, as will shortly be shown to be the case, this distinction determines whether agents will remain or exit if their first market interaction is a tie. This, in turn, determines the equilibrium inflow and outflow patterns that characterize the equilibrium.

Lemma III.1. *In any steady state equilibrium with entry, it must be that* $EU_0 \equiv E_{t=0}[U_i] = 0$.

Proof. This is an accounting exercise. Since ex-ante all participants are identical and any interaction between any two participants is zero sum, no agent can expect to make positive surplus. □

That in any equilibrium with entry agents are indifferent about entering at time $t = 0$ is important because it implicitly defines the threshold likelihood that agents use for all entry/exit decisions, not just the initial one.

Lemma III.2. *The likelihood* $\Lambda(T)$ *is sufficient for the entry/exit decision at time* T*. Furthermore in a steady state equilibrium with entry, it is optimal for each agent to exit at time* T *if and only if* $\Lambda(T) < \lambda_0$.

Proof. The sufficiency follows from the fact that future payoffs depend only on one's expected type and not on the specific history of signals by which this expectation was constructed. That a threshold policy is optimal is a standard result for this stationary problem, and that λ_0 is the threshold likelihood follows from Lemma III.1. □

Now, let us consider the two cases in which ties are either good or bad news in more detail. Since the steady state measures H_s and L_s are equilibrium objects, the approach is to conjecture an equilibrium in which, say $H_s > L_s$, and then infer what this implies about parameters H, L, and δ. I will show that this approach partitions the parameter space.

Case 1: When Ties Induce Optimism, $H_s > L_s$

Conjecture an equilibrium with entry in which $H_s > L_s$. Let $\rho = \frac{H}{H+L}$ and $\rho_s = \frac{H_s}{H_s+L_s}$ be the proportion of high types in the inflow and in the steady state pool, respectively. As previously argued, agents only voluntarily exit after a loss; a win or a tie both increase an agent's likelihood ratio Λ. Since high types never lose, in steady state

$$H = \delta H_s \tag{3.3}$$

The left hand side is the inflow and the right hand side the outflow, with high type agents leaving only exogenously. Low types enter and choose to exit only upon their first loss. The steady state condition for low types is

$$L = (\delta + (1-\delta)\rho_s)L_s \tag{3.4}$$

The right hand side describes the outflow and is comprised of those that exit exogenously and of those who do not but who meet a high type and lose. These two equations then imply a unique steady state H_s and L_s. Given this steady state, it must be verified that entry in period zero is optimal.

$$\begin{aligned} EU_0 &= x\rho\left((1-\rho_s) + (1-\delta)(1-\rho_s) + (1-\delta)^2(1-\rho_s) + ...\right) \\ &\quad - x(1-\rho)\left(\rho_s + (1-\delta)(1-\rho_s)\rho_s + ((1-\delta)(1-\rho_s))^2\rho_s + ..\right) \\ &= x\left(\frac{\rho(1-\rho_s)}{\delta} - \frac{(1-\rho)\rho_s}{1-(1-\delta)(1-\rho_s)}\right) \end{aligned} \tag{3.5}$$

Above, an agent with a high type receives a payoff x in every period in which she participates and meets a low type. A low type will lose x at most once and only if she does not exogenously exit prior to meeting a high type.

Lemma III.3. *There exists an equilibrium with steady state entry in which $H_s > L_s$.*

Proof. Follows from plugging (3.3) and (3.4) into (3.5) and obtaining $EU_0 = 0$. This implies not only that the initial decision is a best response but also that the threshold likelihood is $\Lambda^* = \lambda_0$, which rationalizes the decision to remain in the pool conditional on observing a tie. $\square$

Lastly, one must identify the set of exogenous parameters L, H, and δ for which $H_s > L_s$. Using equations (3.3) and (3.4), obtain

$$\frac{H_s}{L_s} > 1 \quad \Leftrightarrow \frac{2\delta H}{-H + \delta L + \sqrt{4\delta^2 HL + (H-\delta L)^2}} > 1 \tag{3.6}$$

By inspection $\frac{H_s}{L_s} \geq 0$, is increasing in H, decreasing in L, and decreasing in δ, all monotonically. Hence there is a plane in $H \times L \times \delta$ space that separates the set for which inequality (3.6) holds from the set for which it does not hold.

Case 2: When Ties Induce Pessimism, $H_s < L_s$

Next conjecture an equilibrium with entry in which $H_s < L_s$. An agent's likelihood ratio now decreases after having observed a tie, forcing an agent to exit after her first trade unless she wins. Taking this behavior as given, the steady state condition for high types is

$$H = H(\rho_s + (1 - \rho_s)\delta) + \delta(H_s - H) \tag{3.7}$$

There are two kinds of high types in the pool: the newest cohort and the rest. Members of the newest cohort can exit either if they meet another high type or exogenously. The rest of the high types, by virtue of having won in their first trade, have discovered their type and will remain until exiting exogenously. Low types either tie or lose upon entry, hence every low type that enters immediately leaves the next period. Thus

$$L = L_s \tag{3.8}$$

With the steady state thus defined by equations (3.7) and (3.8), once again it needs to be verified that initial entry is a best response.

$$\begin{aligned} EU_0 &= x\rho((1 - \rho_s) + (1 - \rho_s)^2((1 - \delta) + (1 - \delta)^2 + ...)) - x(1 - \rho)\rho_s \\ &= x\left(\rho(1 - \rho_s)\left(1 + \frac{(1 - \delta)(1 - \rho_s)}{\delta}\right) - (1 - \rho)\rho_s\right) \end{aligned} \tag{3.9}$$

When an agent is a high type, if her first trade is a win then she stays in the pool until she exits exogenously, otherwise if her first trade is a tie she exits and receive a payoff of zero. When an agent is a low type, she trades only for one period and only loses if she meets a high type.

Lemma III.4. *There exists an equilibrium with steady state entry in which $H_s < L_s$.*

Proof. As in Lemma III.3, plugging (3.7) and (3.8) into (3.9) and obtaining $EU_0 = 0$ suffices. $\square$

Using equations (3.3) and (3.4), the set of parameters for which $H_s < L_s$ is given by the condition

$$\frac{H_s}{L_s} < 1 \quad \Leftrightarrow \quad \frac{H - L + \sqrt{H^2 - 2HL + \frac{4HL}{\delta} + L^2}}{2L} < 1 \tag{3.10}$$

In this parameter space, it is still the case that $\frac{H_s}{L_s} \geq 0$, is increasing in H, decreasing in L, and decreasing in δ, all monotonically. Hence equation (3.10) is also a hyperplane that partitions the parameter space. This leads to the section's main proposition.

Proposition III.1. *For all parameters $L > 0$, $H > 0$, and $0 < \delta < 1$, there exists a steady state equilibrium with entry. Furthermore, the steady state is generically described either by equations (3.3) and (3.4) or (3.7) and (3.8).*

Proof. First, algebraic manipulation reveals that the left hand side of both (3.6) and (3.10) is the same hyperplane, thus except for the hyperplane itself, every point in the parameter space satisfies exactly one of these inequalities. Lemmas III.3 and III.4 ensure the existence of a steady state equilibrium for parameters satisfying (3.6) and (3.10), respectively. For parameters on the hyperplane, both types of equilibrium can be supported since these parameters induce $H_s = L_s$, rationalizing either staying or exiting after the first tie. □

This result demonstrates that a wholly speculative market can emerge in a setting in which agents have no productive reason to interact. Since it is specifically the benefit of information that induces an agent to enter an adversely selected pool, it is important for the robustness of the result that the process by which the agent obtains information is endogenous to the model. That is, the rate at which the agent learns his type in equilibrium depends on the parameters of the problem, yet regardless of the parameters an equilibrium with entry is supported.

It is well known that dynamic models with endogenous learning are notoriously intractable and to alleviate this the current model imposes a structure with discrete types and deterministic outcomes of agents' interactions. While this provides the requisite tractability, the discreteness of the problem hamstrings the study of some comparative statics. In particular, an important question is how do gains from trade affect the trading volume? As it turns out, the current model is not very well suited for such a question.

To see why, imagine for example that $H_s < L_s$ so that ties are bad news. Imagine also that the surplus from the meeting of two agents is some $\varepsilon > 0$. If this surplus ε is small, it will not be enough to convince an agent whose first observation is a tie to remain in the pool. Hence, when gains from trade are small they have no effect on agents' behavior and consequently on the equilibrium trading volume. Once ε is large enough, an agent observing her first tie will be convinced to remain in the pool and thus the composition of the steady state will jump discretely. However, the analysis of this threshold value of ε is quite cumbersome.

I instead amend the original model to allow for a continuum of types, thus resolving the problem of discrete jumps in the policy function of the agents. However, in order to keep the model tractable I must simplify the learning rule.

3.3 Model with a Continuum of Types

The model in this section departs from the one in the previous section in three ways. First, now the potential pool of entrants has measure normalized to one and each agent draws her type from continuous distribution $F(\theta)$. Second, conditional on entering, agents learn their type perfectly after the first interaction regardless of the outcome of that interaction. In this sense information acquisition is exogenous. And third, I build into the model the possibility of gains from trade. Let the payoffs from the meeting of two agents be

$$u_a(a,b) = \begin{cases} \frac{1}{2}\varepsilon & a = b \\ x + \frac{1}{2}\varepsilon & a > b \\ -x + \frac{1}{2}\varepsilon & a < b \end{cases} \tag{3.11}$$

in which ε is the gain from trade between any two agents. In order to ensure that in any match the loser gets a negative payoff, assume that $0 \leq \frac{1}{2}\varepsilon < x$. Note that for $\varepsilon = 0$ the payoffs are identical to that in equation (3.1).

3.3.1 Steady State Equilibrium

Again I conjecture a steady state with entry, follow the implications of this conjecture, and at the end confirm that entry in fact is optimal. Let $F^s(\theta)$ be the steady state mass with a type below θ.[1] There is a threshold trader $\hat{\theta}$ who, when informed of her type, is indifferent between exiting or

[1] Technically F^s is not a probability distribution because the pool accumulates a measure of agents greater than one.

continuing. That is, $\hat{\theta}$ solves

$$
\begin{aligned}
0 &= \int_\theta u_{\hat{\theta}}(\hat{\theta}, \theta)\ dF^s(\theta) \\
&= (x + \frac{1}{2}\varepsilon)F^s(\hat{\theta}) + (-x + \frac{1}{2}\varepsilon)(1 - F^s(\hat{\theta})) \\
F^s(\hat{\theta}) &= \frac{x - \frac{1}{2}\varepsilon}{2x} \equiv \hat{F}^s
\end{aligned}
\tag{3.12}
$$

This identifies the proportion of the time that the threshold type needs to win in the steady state pool in order to be indifferent. For instance, when $\varepsilon = 0$ and there are no gains from trade the threshold type needs to win exactly half of the time. As gains are introduced, the threshold type can tolerate winning less often.

Next, if $\hat{\theta}$ represents the threshold it must be that all types below exit after the first round and all types above remain until exiting exogenously. The steady state conditions are that

$$
F^s(\theta) = \begin{cases} F(\theta) & \theta \leq \hat{\theta} \\ \frac{1}{\delta}(F(\theta) - F(\hat{\theta})) + F(\hat{\theta}) & \theta > \hat{\theta} \end{cases}
\tag{3.13}
$$

The total mass in the steady state pool is thus

$$
M \equiv F(\hat{\theta}) + \frac{1}{\delta}(1 - F(\hat{\theta}))
\tag{3.14}
$$

Threshold type $\hat{\theta}$ wins only when matched with new entrants, which yields

$$
\hat{F}^s = \frac{F(\hat{\theta})}{F(\hat{\theta}) + \frac{1}{\delta}(1 - F(\hat{\theta}))}
$$

Plugging in equation (3.12) and rearranging, the equilibrium threshold type $\hat{\theta}$ is identified as a function of the parameters of the problem.

$$
F(\hat{\theta}) = \frac{x - \frac{1}{2}\varepsilon}{x - \frac{1}{2}\varepsilon + \delta(x + \frac{1}{2}\varepsilon)}
\tag{3.15}
$$

Note by inspection that $\frac{\partial F(\hat{\theta})}{\partial \varepsilon} < 0$, thus gains from trade induce lower types to remain in the pool. The threshold type is sufficient to describe the equilibrium, and now as in the previous section we confirm that conditional on $\hat{\theta}$ from (3.15) initial entry is a best response.

Proposition III.2. *For any $\varepsilon \geq 0$ and any distribution F, there exists a steady state equilibrium with entry characterized by threshold type $\hat{\theta}$ from equation (3.15).*

The proof demonstrates that $EU_0 \geq 0$ for all $\varepsilon \geq 0$. See Appendix A for details. With Proposition III.2 verifying that equation (3.15) in fact characterizes an equilibrium, I explore the comparative static for which this model was developed. Specifically, the question is not simply whether trading volume grows with gains from trade but whether it does so disproportionately. Intuitively, including gains from trade has the direct effect of inducing more traders to remain in the pool keeping fixed the steady state distribution. There is also an indirect effect, in that lower types remaining in the pool begets more lower types doing the same. I can now formally show that this is the case, and that adding gains from trade involves a multiplier effect.

Proposition III.3. (Speculative Multiplier) *Trading volume is increasing and convex in gains from trade. That is,*

$$\frac{dM}{d\varepsilon} > 0 \quad \textit{and} \quad \frac{d^2M}{d\varepsilon^2} > 0.$$

Proof. Follows directly from equations (3.14) and (3.15). □

3.4 Discussion

This paper provides a bridge between the empirical observation that many trading markets are speculative and the theoretical result that rational speculators should not transact with one another. The insight is that aside from the current period speculative payoff, participating in a trade gives an agent information about her ability and this information has value in the continuation game. An agent that considers entry must weigh the expected informational benefits of the first trade against the expected cost of entering an adversely selected pool. I demonstrate that even when the informational content of a trade is endogenous to the equilibrium, in a purely speculative environment a steady state with entry is supported for all parameter values. I also show that as surplus is introduced into the market, it increases trading volume at an increasing rate.

In Proposition III.1, the existence of an equilibrium with entry is obtained for a specific endogenous learning rule. A different exogenous learning rule still yields the existence of such an equilibrium in the continuous model. I conjecture that with any information structure in which traders statistically learn from trades, an equilibrium with entry can be supported with no gains from trade, and demonstrating this formally would be a valuable addition to the analysis.

Another important extension is to include leverage into the model. Agents that can publicly show their ability through previous earnings are able to cheaply borrow money to finance more trading. Whether the existence result still obtains with leverage is unclear. On one hand, when higher types can leverage the market for a potential entrant is more adversely selected. On the other hand, the payoff to finding out that one is a high type also increases. These two effects cancel out in the models presented in this paper and I conjecture that this will continue to be the case when leverage is introduced.

On the empirical front, the model predicts that older cohorts on average should have higher types. An agent remains in the market only if her likelihood has not dipped below the threshold in the past, and the longer the past the more difficult it is to achieve this. Anecdotally, this seems to match the makeup of many trading floors on which the older traders are on average more highly skilled than the younger ones. Along the same lines, the model makes a prediction about hazard rates. Younger cohorts hold beliefs that have higher variance, thus they are more likely to exit in the future than are older cohorts who are quite certain of their ability. This also is true on trading floors, where a trader with one year of experience is much less likely to still be on the job two years later than a trader with ten years experience. Finally, a sharper prediction of the model is that the frequency of losses should predict exit beyond the dollar amount. If the learning story is true, an agent that flips a coin once and loses \$10 should be less likely to exit than one that lost ten coin flips in a row for \$1 each. Whether empirically this is the case in financial markets is something that remains to be tested.

3.5 Appendix

Appendix A: Proof of Existence of a Steady State Equilibrium with Entry for a Continuum of Types

Proposition III.2 *For any $\varepsilon \geq 0$ and any distribution F, there exists a steady state equilibrium with entry characterized by threshold type $\hat{\theta}$ from equation (3.15).*

Proof. It needs to be demonstrated that $EU_0 \geq 0$. An agent anticipates that if her type is below $\hat{\theta}$ she will trade only once and if her type is above $\hat{\theta}$ she will continue to trade until exiting exogenously.

Her ex-ante expected utility is

$$\begin{aligned} EU_0 &= \int_{\theta<\hat{\theta}} \left(\frac{F^s(\theta)}{M} \left(x + \frac{1}{2}\varepsilon \right) + \frac{1 - F^s(\theta}{M} \left(-x + \frac{1}{2}\varepsilon \right) \right) dF(\theta) \\ &+ \frac{1}{\delta} \int_{\theta\geq\hat{\theta}} \left(\frac{F^s(\theta)}{M} \left(x + \frac{1}{2}\varepsilon \right) + \frac{1 - F^s(\theta)}{M} \left(-x + \frac{1}{2}\varepsilon \right) \right) dF(\theta) \end{aligned}$$

Here, $\frac{F^s(\theta)}{M}$ is the probability density in the pool and probability of a tie has zero measure and is omitted. Plugging equations (3.13) and (3.14) into the above and simplifying obtains

$$EU_0 = \frac{\varepsilon x}{\delta \left(\frac{1}{2}\varepsilon + x \right) + \left(x - \frac{1}{2}\varepsilon \right)}$$

Since by assumption $x \geq \frac{1}{2}\varepsilon$, the expression above is positive and exactly equals zero when $\varepsilon = 0$. Hence, an equilibrium with entry can be supported even in the case of no gains from trade. $\square$

BIBLIOGRAPHY

Aumann, R. J. (1976): "Agreeing to Disagree," *The Annals of Statistics*, 4(6), pp. 1236–1239.

Benabou, R., and R. Gertner (1993): "Search with Learning from Prices: Does Increased Inflationary Uncertainty Lead to Higher Markups," *The Review of Economic Studies*, 60(1), pp. 69–93.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), pp. 841–890.

Buhler, S., and D. L. Gartner (2009): "Making Sense of Non-Binding Retail-Price Recommendations," Working Papers 0902, University of Zurich, Socioeconomic Institute.

Chakravarty, S., and T. R. Kaplan (2006): "Manna from Heaven or Forty Years in the Desert: Optimal Allocation without Transfer Payments," *SSRN eLibrary*.

Crawford, V. P., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50(6), pp. 1431–1451.

Diamond, P. A. (1971): "A model of price adjustment," *Journal of Economic Theory*, 3(2), 156 – 168.

Faber, R. P., and M. C. Janssen (2008): "On the Effects of Suggested Prices in Gasoline Markets," Tinbergen Institute Discussion Papers 08-116/1, Tinbergen Institute.

Fullerton, R. L., and R. P. McAfee (1999): "Auctioning Entry into Tournaments," *Journal of Political Economy*, 107(3), pp. 573–605.

Gill, D., and J. E. Thanassoulis (2010): "The Optimal Marketing Mix of Posted Prices, Discounts and Bargaining," Economics Series Working Papers 479, University of Oxford, Department of Economics.

Hartline, J. D., and T. Roughgarden (2008): "Optimal mechanism design and money burning," in *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pp. 75–84, New York, NY, USA. ACM.

Horowitz, J. L. (1992): "The Role of the List Price in Housing Markets: Theory and an Econometric Model," *Journal of Applied Econometrics*, 7(2), pp. 115–129.

Kreps, D. M. (1977): "A note on fulfilled expectations; equilibria," *Journal of Economic Theory*, 14(1), 32–43.

Kyle, A. S. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53(6), pp. 1315–1335.

Lazear, E. P., and S. Rosen (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5), pp. 841–864.

Mathewson, G. F., and R. Winter (1984): "An Economic Theory of Vertical Restraints," *The RAND Journal of Economics*, 15(1), pp. 27–38.

McCall, J. J. (1970): "Economics of Information and Job Search," *The Quarterly Journal of Economics*, 84(1), pp. 113–126.

Milgrom, and Stokey (1982): "Information, trade and common knowledge," *J. Econ. Theory*, 26, 17–27.

Moldovanu, B., and A. Sela (2001): "The Optimal Allocation of Prizes in Contests," *The American Economic Review*, 91(3), pp. 542–558.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6(1), 58–73.

Odean, T. (1999): "Do Investors Trade Too Much?," *American Economic Review*, 89(5), 1279–1298.

Puppe, C., and S. Rosenkranz (2006): "Why Suggest Non-Binding Retail Prices?," Working Papers 06-10, Utrecht School of Economics.

Reinganum, J. F. (1979): "A Simple Model of Equilibrium Price Dispersion," *The Journal of Political Economy*, 87(4), pp. 851–858.

Rothschild, M. (1974): "Searching for the Lowest Price When the Distribution of Prices Is Unknown," *The Journal of Political Economy*, 82(4), pp. 689–711.

Rothschild, M., and J. Stiglitz (1970): "Increasing risk: I. A definition," *Journal of Economic Theory*, 2(3), 225–243.

Spence, M. (1973): "Job Market Signaling," *The Quarterly Journal of Economics*, 87(3), 355–374.

Tirole, J. (1982): "On the Possibility of Speculation under Rational Expectations," *Econometrica*, 50(5), pp. 1163–1181.

Yang, H., and L. Ye (2008): "Search with learning: understanding asymmetric price adjustments," *RAND Journal of Economics*, 39(2), 547–564.