Screening in Vertical Oligopolies

Hector Chade* and Jeroen Swinkels†

January 2019

Abstract

A finite number of vertically differentiated firms simultaneously compete for and screen a continuum of agents with private information about their ability or their willingness to pay for quality. Firms compete through menus of wage-effort or transfer-quality pairs. In equilibrium, higher firms serve higher segments of types. In each segment, the allocation is distorted downward from the efficient level on types below a threshold, but upwards above. The equilibrium approaches the competitive limit as the number of firms grows large. The welfare effects of private information may be reversed from the monopoly setting. While payoffs in this game are neither quasi-concave nor continuous, we show that if firms are sufficiently differentiated, then any strategy profile that satisfies a simple set of necessary conditions is an equilibrium, and we show that an equilibrium exists.

Keywords. Adverse Selection, Screening, Quality Distortions, Oligopoly, Incentive Compatibility, Positive Assortative Matching, Vertical Differentiation.

*Arizona State University, hector.chade@asu.edu

†Northwestern University, j-swinkels@northwestern.edu

We thank Meghan Busse, Laura Doval, Bruno Jullien, Andreas Kleiner, Natalia Kovrijnykh, Michael Powell, Luis Rayo, and especially John Conlon, along with seminar participants at Northwestern, 2018 SED–Mexico City, Universidad Alberto Hurtado–Chile, CIDE-Mexico, University of Georgia, Universitat Pompeu Fabra, and Universidad Nacional de Cuyo-Argentina, for their helpful comments.

1 Introduction

Screening is an important feature of labor and product markets, and is a pervasive topic in economic theory and applied work. Indeed, the principal-agent model with adverse selection developed in Mussa and Rosen (1978) and Maskin and Riley (1984)–where a monopolist screens a consumer who has private information about his valuation by providing different quantities or qualities of a good at different prices–is a workhorse in the economics of information. At the other extreme is the competitive screening model of Rothschild and Stiglitz (1976) and variations thereof–where identical insurance companies screen consumers who differ in and have private information about their riskiness.

Many markets with screening do not fall at these extremes. Instead, a small number of heterogeneous firms both screen their own customers and compete for them in the first place. Examples abound in health care, labor, and product markets. For example, luxury handbag manufacturer Saint Laurent must screen among its customers. That is, as it chooses the quality and price of any given handbag in its line-up, it must consider the effects of these choices on the sales of its other handbags. But, unlike a monopolist, these choices by Saint Laurent also affect how successfully they compete with Hermès above them and Coach below. Indeed, almost all consumer-packaged-goods firms sell multiple products at different quality and price points, but do so in an environment with heterogeneous competitors. Similarly, firms that compete for talent often face the problem of screening their workers into appropriate roles, but also face competition with vertically differentiated rivals.

Although there are some notable attempts in the literature (see the review below), the lack of a standard workhorse for this case has hindered progress. This is true both for theoretical work–where oligopolistic screening is not well understood–and for recent empirical work–which estimates models of insurance and product markets with an oligopolistic structure but where informational frictions and screening take a restrictive or reduced form.

This paper is an attempt to fill this gap. We develop a natural extension of the Mussa and Rosen (1978) and Maskin and Riley (1984) principal-agent paradigm to an oligopolistic setup with a finite set of vertically differentiated firms and a continuum of agents with private information. For definiteness, we cast the analysis as a labor market, where workers have private information about their ability, but the model can be reinterpreted as a product market where customers have private information about their willingness to pay for quality or quantity (with exclusive dealing).

We provide necessary conditions that equilibria must satisfy and then show that, under some economically interpretable assumptions, these conditions are sufficient, a result that also allows us to prove equilibrium existence. We shed light on the properties of the competitive limit of the model as the number of firms grows large, and study the welfare effects of asymmetric information.

In the model, firms differ in the technology by which they transform a worker's effort into revenue. Consistent with vertical differentiation, these technologies are ordered by a single-crossing property, so firms with higher index have a higher marginal revenue for effort.[1] The technology of each firm is additive over all the workers hired (no peer effects). On the other side of the market there is a continuum of workers with quasilinear preferences who differ in their marginal disutility of effort. Ability is a worker's private information, and can take on a continuum of values.

Competition is modelled as the following two-stage game. In stage one, firms simultaneously post menus of contracts, where a menu consists of wage-effort pairs, or, equivalently, utility-effort pairs, one for each type, and where these menus can be restricted to be incentive compatible. Thus, by assumption, we rule out contracts that, for example, condition on the offers of other firms. This is not without loss of generality, but both adds tractability and seems the economically reasonable assumption in most settings.[2] In stage two, each worker chooses the firm and the contract from its menu that suits him best, resolving ties across firms equiprobably. By folding the workers' sequentially rational behavior into the model, we analyze the problem as a simultaneous game among the firms and then study the properties of its pure strategy Nash equilibria. A challenge we face is that this game has an infinite dimensional strategy space and discontinuous payoffs.

We first derive a set of properties that any equilibrium exhibits. Several of these necessary conditions are closely related to ones that Jullien (2000) derives in the case of single principal who faces a type-dependent participation constraint, where in our setting, this constraint is determined by the most attractive contract the worker of any given ability faces from an alternative employer. We provide alternative proofs for these conditions because some details of our setting are different (in particular at ties) and because we think there is value in elementary proofs that avoid optimal control tools. We will discuss this connection in detail below.

Since our model has private values–the type of a worker enters the firm's profit only through the contract chosen by the worker–we show that in equilibrium firms make *positive profit* on each worker they hire. Any equilibrium also satisfies *no poaching*: if a firm does not hire a type, then imitating the menu offered by the incumbent to that type yields negative profit to the imitating firm. An implication is that the worker is matched to the firm that most efficiently uses the effort level he exerts. This does not imply that the equilibrium match is efficient: more total surplus may be generated if the worker were matched to a different firm at a different effort level.

Since our model embeds a nontrivial matching problem between firms and workers, an important task is to study equilibrium sorting patterns. We show that any equilibrium entails *positive sorting*: firms with a higher index hire intervals of workers with higher types. For each firm, there is a region of types that are hired exclusively by that firm. If firms are not very differentiated,

[1] A model with both vertical and horizontal differentiation would also be of great interest, but is beyond the scope of this paper.

[2] For more on general mechanisms, see Epstein and Peters (1999) and Martimort and Stole (2002). In particular by Corollary 1 in Martimort and Stole (2002), there is no further loss of generality in assuming the firms simply post menus as they do here. See also the "Extended Example" in Martimort and Stole (2002), Section 5.

then adjacent firms may also tie on an interval of types at the boundary between successive firms. If they do so, then they offer a pooling contract to those types and competition drives profits on those types to zero.

Equilibrium sorting highlights the dual role that menus play in our model: they are used to screen internally the types of the workers hired, and they also serve to attract the right pool of types for the firm. Positive sorting is straightforward in the complete information version of our model due to the supermodularity assumptions and the absence of peer effects within firms. It is more subtle under incomplete information due to the incentive compatibility constraints.

Positive sorting proves fundamental in the analysis of the properties of equilibrium menus and distortions. Since firms hire intervals of types, we can focus on each firm solving for the optimal endpoints of the interval hired, and the optimal menu given those endpoints, subject to the endogenous type-dependent participation constraints induced by the menus offered by the other firms. On the interval of types for which any given firm is the uniquely best choice in equilibrium, the menu offered by the firm satisfies *internal optimality*–the effort level of each type is pinned down by a condition that generalizes the standard trade off between efficiency and information rents. This generalization reflects both that the firm serves only a segment of the market and that it is in general competing with firms both below and above them in the marketplace, and so faces more than one binding participation constraint. In addition, each firm must satisfy *optimal boundary* conditions that determine the endpoints of the relevant interval of types. These conditions reflect the trade off that changing the effort given to a boundary type alters the profits earned on that type, and also attracts or loses some marginal types.

The internal optimality and optimal boundary conditions yield a clear-cut pattern of equilibrium distortions. The highest firm distorts effort *downwards* for all types. This reflects the standard intuition that lowering the effort of any given type employed flattens the surplus function and hence lowers the information rents of higher types. In turn, the lowest firm distorts effort *upwards* for all types. This follows because for the lowest firm, the outside option (of working for someone else) binds only for the highest type of worker hired. Raising the effort level of any given type thus steepens the surplus function, which lowers the information rents of all lower types.

Consider next a middle firm. Here, the key is that, holding fixed the utility of the lowest and highest worker the firm employs–the two types on which the participation constraint will turn out to bind–the firm can lower the information rents of workers in the middle of the interval employed by simultaneously lowering the effort of low types employed and raising the effort of high types employed. Doing so effectively makes surplus more convex, and, holding fixed the utility at the endpoints, this lowers the information rents of middle types. Hence, the firm distorts effort *downwards* on workers below a threshold, and *upwards* for those above.

In an interesting precedent dating to 1849, Dupuit (1962) observes that a rail company provides roofless carriages in third class to "frighten the rich." As confirmed by the extensive literature on screening, the low quality third-class car helps to sell second class seats at a higher price. But Dupuit goes further. The first-class passenger receives a "superfluous" level of quality. In the standard model, this is not so: all quality distortions are downward. But, in our model, if the rail firm is a middle firm that competes for its richest customers against, for example, a private carriage, then the high type served does indeed receive an inefficiently high level of service. From their point of view, the extra quality is nearly worth the extra cost, and so the extra cost can be largely reflected in the price. But, the superfluous extra quality–with the extra price implied–reduces the temptation for the second-class passenger to ride first, and this also helps to reduce the information rents of the second-class passenger.

Another implication of the analysis is that, when firms are sufficiently differentiated, there are effort gaps at the boundaries between adjacent firms. This is testable: in our setting, we should observe better firms having strictly more productive workers. Similarly, we should observe that products of certain intermediate qualities are simply not offered in some markets.

Next, we study the behavior of our oligopoly model as the number of firms grows large and the vertical differentiation between successive firms becomes small. We show that each firm's profit is bounded above by a constant that goes to zero, as in the competitive limit. In turn, the utility that each type receives in equilibrium converges to the surplus in the efficient match and effort level, again as in the competitive limit.

We finish our examination of the implications of necessity by comparing the outcome of the oligopoly market with and without asymmetric information. In a monopoly, it is unambiguous that full information hurts the worker by destroying information rents, and helps the firm, which gains back the information rents, and further can now induce efficient effort. Here, we have a surprising reversal, driven by the new force that under full information a firm can now add a wage-utility pair designed to poach the workers of another firm without worrying about potentially attracting their own existing workers. This competitive force dominates at least for types where the second most efficient firm for the type is nearly as efficient as the first, and thus full information *helps* the worker and *hurts* the firm.

We then turn to the analysis of sufficient conditions for an equilibrium and to the question of equilibrium existence. We show that if firms are differentiated enough–a condition we call *stacking*–then first, any strategy profile that satisfies positive assortative matching plus the internal optimality and optimal boundaries conditions is (essentially) an equilibrium; and, second, an equilibrium *exists*. We view these results as a central contribution of the paper, since first, they are the most challenging technical problems that one must tackle in this setting, second, they are fundamental for economic applications of our model, and third, they are novel in the literature.

The technical challenges in these results derive from the fact that this is a game whose payoff functions are neither continuous nor quasi-concave in the strategy profile. The failure of continuity comes from standard tie-breaking considerations: A firm that is offering a little less surplus than its competitors never wins, while one that offers a little more does so always. The failure of quasi-concavity comes from the phenomenon that two strategies for a given player may earn the same payoff, but hire different sets of workers. Because of this, a convex combination of these two strategies will typically hire a different set of workers than either of the two, and so have payoff that does not relate to either of them in a tractable fashion.

The property that payoffs are not quasi-concave in the strategy chosen makes the fact that our necessary conditions are also sufficient both surprising and non-trivial. In addition, the lack of quasi-concavity also makes convexity of each player's best responses non-obvious, which complicates the use of off-the-shelf existence results.

The first step of our attack is to restrict attention to menus that yield non-negative profits on all types, and for which the induced effort levels are within certain bounds that our necessary conditions suggest are reasonable. We show that under stacking, if other firms choose menus with these two properties, then each firm will best respond with menus that also satisfy them. More importantly, stacking then implies that the resulting strategy profile has positive sorting and that the optimal boundary conditions imply that no poaching holds as well, which deals with inframarginal types.

The second step is to effectively reduce the problem of finding a best response to two dimensions. We show that under the stated properties, each firm can restrict attention to choosing the upper and lower boundaries of the interval of types hired, with the rest of the menu pinned down by internal optimality. Using stacking, we show that payoffs are continuous in this parameterization. A direct implication is that the set of best responses is nonempty for each firm.

Even viewed as a two-dimensional optimization problem, we still face significant technical challenges. Our third key step is an exercise in topography. Fix the behavior of a firm's opponents, and consider a landscape given by the payoff to the firm, where the choice of the bottom endpoint is a choice from west to east, while the choice of the top endpoint is a choice from south to north.

This landscape has valleys, and indeed, local minima. But, when the other firms play strategies from the previously defined set, we show that the firm has available positive profit strategies. So, consider the "islands" where the payoff is positive. These islands have terrain that is kinked, because the participation constraint–given by the utility offered to each type by the toughest competitor of the firm–has kinks at each type where the relevant opponent changes. Nor need payoffs be quasi-concave even on a given island.

We show first that on such islands, any place where the first order conditions are satisfied is also a local maximum, so that any local minima are underwater. Next, fix any position from south to north (a choice of the top endpoint) such that there is some land at that latitude. We show that this defines a single interval of latitudes.

Fix some such latitude, and consider moving west to east. We show that there is a single interval where one is above water, and that payoffs are strictly quasi-concave as one moves from west to east in this interval. Hence there is a unique highest point at each latitude. But then, there is a single island, and there is a unique path running from the south the north end of the island with the property that each point along the path is the highest point at that latitude. Finally, we show that payoffs are strictly quasi-concave as one moves northward along this ridge. It follows that the island has a unique peak, and that any point that satisfies the first-order conditions–the optimal boundary conditions–along with the positive profit condition, is in fact that maximum.[3]

Our sufficiency result follows from these steps. Under stacking, if a strategy profile satisfies positive sorting, internal optimality, and the optimal boundary conditions, then it pins down uniquely the interval of types hired by each firm and the optimal effort function. The endpoints of this interval are a local maximum in the reparameterized problem by the optimal boundary condition. But then, since the reparameterized problem has a unique optimum, each player is in fact best-responding. And outside the interval of types hired by each firm, one can easily modify the menu to comply with the bounds on effort and positive profits assumed, completing the construction of an equilibrium strategy profile.

It remains to show that an equilibrium exists. To this end, we further restrict attention to a class of strategies that also satisfies a natural bound on the slope of the effort function offered by each firm, and a natural lower bound on the utility that is offered. We show that if other firms use strategies that satisfy these conditions, then each firm has a best response with the same properties. This class of strategies is sufficiently well-behaved to permit the application of a standard fixed point theorem, delivering existence. In particular, because we have shown that for any given behavior of its opponents, the firm has a unique optimal interval served, and because internal optimality ties down what is happening on that interval, any two best responses will differ only in inessential ways, and the set of such best responses will be convex.

The next section reviews the literature. Section 3 describes the model. Section 4 derives the necessary conditions and their implications. Section 5 studies quantity discounts, the competitive limit, and the welfare effects of asymmetric information. Section 6 focuses on sufficiency and existence, presenting the main results and describing the main steps of the proofs. Section 7 concludes. Appendices A and B contain all omitted proofs as well as subsidiary lemmas.

2 Related Literature

The paper is clearly related to the huge literature on principal-agent models with adverse selection, as in Mussa and Rosen (1978), and Maskin and Riley (1984), and the host of papers that build on them (see Laffont and Martimort (2002) for a survey). It is more related to the small literature on

[3]The reader may wonder why we did not simply establish that the relevant function is strictly concave at any critical point. First, our function fails this property–there can be local minima "under water." Second, in $\mathbb{R}^2$ this property is not enough. For example (see Chamberland (2015), pp. 106–108), $f(x, y) = -(x^2-1)^2-(x^2y-x-1)^2$ has only two critical points, one at $(-1, 0)$ and one at $(1, 2)$, where both are strict local (indeed global) maxima.

oligopoly and price discrimination under adverse selection, nicely surveyed by Stole (2007). The most relevant related papers are Champsaur and Rochet (1989), Spulber (1989), Stole (1995), and Biglaiser and Mezzetti (1993). Finally, since competition makes each agent's reservation utility endogenous and type-dependent, the paper is related to the general analysis of the principal-agent problem with adverse selection and type-dependent reservation utility in Jullien (2000).

Champsaur and Rochet (1989) analyze a two-stage game where two identical firms choose intervals of qualities they can produce, and then in the second stage offer price schedules to consumers. The two-stage nature of their model gives scope for the firms to cede parts of the market before price competition takes place, giving very different underlying economics.

Spulber (1989), working in a Salop (1979) circular model of horizontal differentiation, considers screening on quantities. The relationship between his analysis and ours is not that close. In particular, his surplus schedule has the same structure as in monopoly, with the intercept determined by the level of competition.

Closer to our paper is Stole (1995), who analyzes an oligopoly setting with screening in which a continuum of customers differ along both a vertical and a horizontal dimension, but only one of them is allowed to be private information, so as to keep the analysis one-dimensional. When the horizontal differentiation parameter is private information, then the model is an extension of Spulber (1989), and so again not that close to ours. The more interesting case for our purposes is when the vertical dimension is private information while the horizontal one is common knowledge, and where the firms can tailor their offerings to the horizontal type of the agent. In this case, and under symmetric marginal costs, when there are two firms the market partitions into two intervals with each firm serving all vertical types of those customers closest to it. Competition leads the distant firm to offer the product at marginal cost (which is assumed constant across quality levels), and the optimal menu makes customers up to a threshold type indifferent between buying from either firm, while above that type the menu is as in the monopoly screening case. With multiple firms located in a circle, the paper shows how entry reduces distortions in quality.

The critical aspect of Stole's analysis is that the cost of providing quality is the same across firms. Because of this, if a customer is closer to Firm 1 than to Firm 2, it is efficient (and the equilibrium outcome) for Firm 1 to serve the customer *regardless of the customer's vertical type.* But then, the close-by firm faces a standard monopoly screening problem with the outside option of the customer determined by the best offer the more distant firm can make without losing money.

In contrast, our model has no horizontal type. But, in our product market interpretation, the cost to providing incremental quality *differs* across firms. It is presumably relatively expensive for Ferrari to produce a basic economy vehicle, and prohibitively expensive for Hyundai to produce a purebred performance automobile. Hence, it is not clear *a priori* to which firm the customer should be assigned. Indeed, how customers are matched to firms is at the heart of our analysis.

Biglaiser and Mezzetti (1993) analyze a model with adverse selection and "false moral hazard" between two heterogeneous principals who differ in their cost of production, and one agent with a continuum of possible types. In equilibrium, the principal with lower marginal cost serves an interval of types higher than a threshold, the other principal dominates the low-type segment of the market, and both offer the same pooling contract for intermediate types, driving profits to zero on that segment. Except for the false moral hazard, this is a special case of our setup, for which we provide a complete equilibrium analysis. Biglaiser and Mezzetti (1993) assume that ties are broken in favor of the firm that gains the most from that type. This tames payoff discontinuities at ties in a crucial way, but is less economically natural than equiprobable tie-breaking.

Another important reference is Jullien (2000). He provides a sophisticated analysis of optimal menus in a principal-agent model with exogenously given type-dependent reservation utility, and shows that both upward and downward distortions can emerge.[4] Holding fixed the behavior of the other firms, the problem facing each of our firms is similar to the one in Jullien (2000), with the key difference being that he assumes that if the firm offers the worker surplus equal to his outside option, then the agent accepts, while in the oligopoly setting, ties are broken equiprobably. This makes some difference at a technical level. Existence of a best-response is no longer guaranteed, and it becomes harder to analyze the problem using standard control techniques. Because of this, we have to work harder than Jullien (2000) to prove, for example, that any optimal contract implements actions that are continuous over the range of types employed.

Because of the similarity of the underlying problems, those of our necessary conditions that derive solely from the implications of best-responses on a firm-by-firm basis are the same as those in Jullien (2000). In particular, our positive profit, internal optimality, and optimal boundary conditions each has a close relative in Jullien. In turn, those of our necessary conditions which are derived from the interplay of the incentives of one firm and another are novel. And since our model with competition endogenizes the agents' type-dependent reservation utility, we can provide more clear-cut predictions of the pattern of distortions that must arise in any equilibrium.

In his Theorem 4, Jullien (2000) shows that under one of two conditions (which our model satisfies) his necessary conditions are also sufficient with full participation, while in Section 4, he describes how to extend his model to also handle cases where full participation is not optimal. It is tempting to conclude that, suitably modified, Jullien's analysis also implies that our necessary conditions are sufficient for optimality in our setting where each firm hires only some of the workers. However, as that paper recognizes (p.17, second paragraph), one must be extremely careful in applying the sufficiency part of Theorem 4 and the ideas of Section 4 at the same time.

To see the issue, the idea of Section 4 in Jullien (2000) is to add an artificial technology that mimics the action and surplus that the agent gets at his outside option (in our setting, the agent's favorite offering from the other firms) but does so at zero profit for the firm. The firm's profits are then the maximum of those associated with the original production function and the artificial

[4]This generalizes the main insight in Maggi and Rodriguez-Clare (1995).

technology. With this modified production function, the principal is willing to hire all workers, and so, Jullien argues, we can solve the full participation problem per his original analysis, and then simply drop workers in the (zero profit) regions where the modification was relevant.

The problem is that the heart of the proof of sufficiency in Theorem 4 (in either variation) is that the benefit to the firm from effort is concave. But, because the modified profit function in Section 4 of Jullien (2000) involves a maximum of two functions, it will typically have an upward kink where it transitions from one function to the other. Hence, the modified production function is not concave, and Jullien's analysis does not apply. One major contribution of our paper is to provide a proof of sufficiency that allows for the fact that with competition, each firm will have less than full participation. This construction is also at the heart of our existence proof.[5]

Since our model embeds a nontrivial matching problem between firms and sets of workers, the paper relates to the literature on many-to-one matching problems with transfers, as in the seminal papers of Crawford and Knoer (1981) and Kelso and Crawford (1982), who provide conditions under which a competitive equilibrium exists without private information. A recent paper that sheds light on sorting in matching models with "large" firms (and complete information) is Eeckhout and Kircher (2018). Our model can be thought of as a matching setting under one-sided incomplete information, where firms that differ in their technology compete in a noncooperative fashion for sets of workers with private information about their disutility of effort.[6]

Finally, there is a large literature on competitive markets with adverse selection in the tradition of Rothschild and Stiglitz (1976), including some recent contributions that feature imperfect competition driven by search frictions, as in Guerrieri, Shimer, and Wright (2010) and Lester, Shourideh, Venkateswaran, and Zetlin-Jones (2018). Our setup differs in two fundamental ways from that literature: first, we have a small number of heterogeneous firms or principals, and thus there is oligopolistic competition and a nontrivial sorting problem; and second, unlike the insurance problem ours is a model with private values. A tighter connection between our approach and this literature must await the extension of our analysis to common values.

3 The Model and Equilibrium

There is a unit measure of agents (workers) and there are N principals (firms). Agents differ in a parameter $\theta \in [\underline{\theta}, \overline{\theta}]$ with cumulative distribution function (cdf) H with strictly positive and $\mathcal{C}^1$

[5] Our results thus also prove sufficiency for a class of type-dependent reservation utilty models not covered by Jullien's analyis. In essence, one needs the slope of the type-dependent reservation utility to satisfy a "steep or shallow" condition similar to our stacking condition. This is natural in our setting if firms are sufficiently vertically differentiated, and may be natural in others.

[6] Another example in the literature with positive sorting under incomplete information can be found in Liu, Mailath, Postlewaite, and Samuelson (2014).

density h.[7] We assume H and $1-H$ are strictly log-concave.[8]

Each worker exerts effort $a \geq 0$ at cost $c(a,\theta) = (1-\theta)a$.[9] If type θ exerts effort a and obtains wage w, then his utility is $w - c(a,\theta)$. For simplicity, we assume that the worker has no outside option beyond choosing among the offers of the various firms. To zero in on competition under adverse selection, we assume that effort is observable, thus ruling out moral hazard issues.

If firm n pays wage w to a worker who exerts effort a, then the firm's payoff is $B^n(a) - w$, where B^n is $\mathcal{C}^2$ and strictly concave in a. We assume that higher indexed firms put higher value on incremental effort, so that $B^n_a > B^{n-1}_a$ for all n, or, equivalently, B^n is supermodular in (a,n). Firms do not have capacity constraints and their technology is additively separable across workers.

Firms simultaneously offer menus of contracts, where Firm n's menu is a pair of functions (α^n, w^n), where $\alpha^n(\theta)$ is the action required of a worker who chooses Firm n and announces type θ, and $w^n(\theta)$ is his wage. Contracts are exclusive: each worker can work for only one firm. As mentioned in the introduction, we rule out contracts that depend on other firms' offers.

Let v^n be the surplus function for a worker who takes the contract of firm n, given by $v^n(\theta) = w^n(\theta) - c(\alpha^n(\theta),\theta)$. It is without loss that firms offer incentive compatible menus. Thus, a menu can equally well be described as (α^n, v^n), where, as is standard, incentive compatibility is equivalent to requiring that the action schedule α^n is increasing and that $v^n(\theta) = v^n(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \alpha^n(\tau)\, d\tau$ for all θ (so in particular, v^n is convex). We will do so henceforth. Let player n's strategy set, S^n, be the set of such pairs $s^n = (\alpha^n, v^n)$. The joint strategy space is $S = S^1 \times \cdots \times S^N$ with typical element s. Let $s^{-n} \in \times_{n' \neq n} S^{n'}$ be a typical strategy profile for players other than n.

The principal's profit on an agent of type θ who takes action a and is given utility v_0 is

$$\pi^n(\theta, a, v_0) = B^n(a) - c(a,\theta) - v_0. \tag{1}$$

For any menu (α, v), we write $\pi^n(\theta, \alpha, v)$ as shorthand for $\pi^n(\theta, \alpha(\theta), v(\theta))$.

After observing the posted menus, workers sort themselves to the most advantageous firm. Formally, for any n, for any $s^{-n} \in S^{-n}$, define the scalar-valued function v^{-n} given by

$$v^{-n}(\theta) = \max_{n' \neq n} v^{n'}(\theta)$$

as the most surplus offered by any of n's competitors. As the maximum of convex functions, v^{-n} is convex. Let a^{-n} be the associated scalar-valued action function, so that a^{-n} is an increasing

[7] We use increasing and decreasing in the weak sense of nondecreasing and nonincreasing, adding 'strictly' when needed, and similarly with positive and negative, and concave and convex. Also, for any function f and argument x of f, we write $(f)_x$ for the total derivative of f with respect to x. We use the symbol $=_s$ to indicate that the objects on either side have strictly the same sign. We follow the hierarchy Lemma, Proposition, Theorem. Finally, wherever it is clear which firm we are talking about, we suppress the n superscript.

[8] As is standard, our model is equivalent to one with a single worker drawn from H.

[9] It natural in this interpretation to assume $[\underline{\theta}, \bar{\theta}] \subseteq [0,1]$, but this plays no formal role. The functional form for the cost function adds tractibility and we believe does not subtract significantly from the economics of the situation.

function and almost everywhere equal to v_θ^{-n}. From the point of view of n, (a^{-n}, v^{-n}) summarizes all the relevant information about the strategy profile of his opponents. Define $\varphi^n(\theta, s)$ as the probability that n hires θ given $s \in S$. We assume that ties are broken equiprobably,[10] so that

$$\varphi^n(\theta, s) = \begin{cases} 0 & v^n(\theta) < v^{-n}(\theta) \\ \frac{1}{\#\{n' \in \{1,\dots,N\} | v^{n'}(\theta) = v^{-n}(\theta)\}} & v^n(\theta) = v^{-n}(\theta) \\ 1 & v^n(\theta) > v^{-n}(\theta) \end{cases}.$$

Define

$$\Pi^n(s) = \int \pi^n(\theta, \alpha^n, v^n)\, \varphi^n(\theta, s)\, h(\theta)\, d\theta$$

as the profits to firm n given strategy profile s. The form of Π^n reflects our assumptions that there are no capacity constraints and the technology is additively separable across workers. Because the optimal behavior of the workers is already embedded in φ (given our tie-breaking assumption), we can view the game as simply one among the firms, with strategy set S^n and payoff function Π^n for each n. Let $BR^n(s) = \arg\max_{s^n \in S^n} \Pi^n(s^n, s^{-n})$. Strategy profile s is a Nash equilibrium (in pure strategies) of $(S^n, \Pi^n)_{n=1}^N$ if for each n, $s^n \in BR^n(s)$.

Define $\alpha_*^n(\theta) = \arg\max_a (B^n(a) - c(a, \theta))$ as the efficient action when θ and n are matched, and let $v_*^n(\theta) = B^n(\alpha_*^n(\theta)) - c(\alpha_*^n(\theta), \theta)$ be the most surplus n can offer type θ without losing money. Define $v_*(\theta) = \max_{n \in \{1,\dots N\}} v_*^n(\theta)$ as the most surplus that any firm can offer type θ, and let $v_*^{-n}(\theta) = \max_{n' \neq n} v_*^{n'}(\theta)$ be the most surplus that any firm other than n can offer θ. We will assume that for all n there exists θ_*^n such that

$$v_*^n(\theta_*^n) > v_*^{-n}(\theta_*^n). \tag{2}$$

That is, there is some type θ_*^n such that n is the unique firm that can create maximal surplus for type θ_*^n. Because $B^n - c$ is strictly supermodular in (a, θ), the sets over which each firm is the unique maximizer of $v_*^n(\theta)$ will be intervals ordered by the names of the firms, and so in particular, we can take $\theta_*^1 = \underline{\theta}$, and $\theta_*^N = \bar{\theta}$. As will be seen, the existence of θ_*^n for all n will imply that each firm in equilibrium employs a positive measure of workers and earn strictly positive profits.

3.1 Interpretation as a Product Market

To reinterpret our model as one of product quality, assume that each customer has demand for a single unit, and that the value of consuming a unit of a product of quality a is $-c(a, \theta) = (\theta - 1)a$, where we take $\underline{\theta} > 1$, and where since $-c_{a\theta} = 1$, higher θ types put higher marginal value on quality. The production cost of quality a to firm n is $-B^n(a)$, so that higher indexed firms have

[10] Equiprobably tie-breaking seems to us economically natural and it is also tractable. Examination of the proofs of Corollary 3 and Proposition 4 in the Appendix, however, reveals that the details of tie-breaking are inessential as long as φ^n is strictly positive wherever $v^n(\theta) = v^{-n}(\theta)$.

lower incremental cost of providing quality.[11] A reinterpretation in terms of quantity provision under exclusive dealing is also straightforward.

4 Necessity

In this section, we present a set of necessary conditions for a Nash equilibrium in pure strategies. For convenience, we will state the main theorem of this section before defining the various terms involved. Then, we define each of the terms, and, for each necessary condition, show why it must hold, and flesh out its economic implications. All proofs are in Appendix A.

Theorem 1 (Necessity) *Every pure strategy Nash equilibrium with no extraneous offers satisfies positive profits on each worker hired, no poaching, positive sorting, internal optimality, and optimal boundaries.*

4.1 Positive Profits on Each Worker Hired (PP)

The *positive profits* condition (PP) is satisfied if for each n, the probability that n hires a worker on whom he strictly loses money is 0. We prove–and use several times in what follows–the stronger statement that for any $s = (s^n, s^{-n})$ (equilibrium or not), s^n can be transformed to a strategy that is equivalent to s^n anywhere s^n earns positive profits, but eliminates any situation where s^n loses money. To see the intuition, let P be the set of types on which s^n makes money. Eliminate all action-wage offerings for workers not in P. Workers in P have fewer deviations available, and so truthful reporting remains incentive compatible for them. Workers not in P who go to another firm save the firm money. Finally, workers not in P who now accept the same contract as some worker in P are now profitable because, using private values, the firm is indifferent about the identity of the worker conditional on his accepting a given offer.

One key implication of PP is that each firm earns strictly positive profits in equilibrium: since other firms do not lose money, a firm that offers the menu $(\alpha_*^n, v_*^n - \varepsilon)$ for ε sufficiently small will win at a minimum near θ_*^n and earn strictly positive profits on any workers hired. See Corollary 3. Another key implication is that there is no cross-subsidization: losing money on some types does not enhance the profits earned on others.

4.2 No Poaching (NP)

The *no poaching condition* (NP) holds if for all θ such that $\varphi^n(\theta, s) < 1$, $\pi^n(\theta, a^{-n}, v^{-n}) \leq 0$. That is, if θ is not always hired by Firm n, then imitating θ's equilibrium contract with the firm that hires θ is unprofitable. Intuitively, if there is an interval of workers where n is not winning always, but can make money by imitating the incumbent, then Firm n can offer those workers a

[11]Note that we did not restrict B^n to be increasing.

deal that replicates that of the incumbent plus some ε in surplus without affecting the behavior of any type on whom it is currently offering uniquely the best deal. As such, NP is about stealing the inframarginal workers of another firm.

Under NP, each firm offers a deal at least as good as could the second most capable firm *at the action chosen*. This bound is strongest when firms have similar capabilities (when π^n and π^{n+1} are similar). The emphasis is important: a firm might profitably outcompete n on type θ with another action, but at the cost of attracting some of its existing workers in a detrimental way.

4.3 Positive Sorting (PS)

To define our positive sorting condition, we need a little notation for what will happen if two firms tie. For $n \in \{1, \ldots, N-1\}$, let $\hat{a}^n$ be the unique solution to $B^n(a) = B^{n+1}(a)$. Let $\hat{v}^n(\cdot) = B^n(\hat{a}^n) - c(\hat{a}^n, \cdot)$. That is, $\hat{a}^n$ is the unique action where firms n and $n+1$ derive the same benefit, and $(\hat{a}^n, \hat{v}^n(\theta))$ earns zero profit on type θ for firms n and $n+1$ given action $\hat{a}^n$.

Say that *positive sorting (PS)* holds for strategy profile s if four things are true: First, for each firm n, there is a single non-empty interval (θ_l^n, θ_h^n) of workers where $\varphi^n(\theta, s) = 1$, so that workers in this interval are always hired by firm n. Second, these intervals are ordered, in that $\theta_h^n \leq \theta_l^{n+1}$ for all n. Third, $\theta_l^1 = \underline{\theta}$, and $\theta_h^N = \bar{\theta}$. Finally, if $\theta_h^n < \theta_l^{n+1}$, then firms n and $n+1$ are offering the same contract $(\hat{a}^n, \hat{v}^n(\theta))$ to each θ in $\left(\theta_h^n, \theta_l^{n+1}\right)$, so that each firm is winning half the time ($\varphi^n = \varphi^{n+1} = 1/2$) and profits are zero on these workers.

An implication of the definition is that $v^n(\theta_l^n) = v^{-n}(\theta_l^n)$ and $v^n(\theta_h^n) = v^{-n}(\theta_h^n)$, where "=" is relaxed to "$\geq$" at $\theta_l^1 = \underline{\theta}$ and $\theta_h^N = \bar{\theta}$.

Proposition 1 *Any pure strategy equilibrium with no extraneous offers has PS.*

To see the intuition for PS, fix $\theta' > \theta$. By incentive compatibility, θ' is taking an action at least as high as θ in equilibrium. But, $B^n(a)$ is strictly supermodular in n and a. Hence, if n sometimes hires θ' and $n' > n$ sometimes hires θ, then, by PP and NP, either n will want to always hire θ, or n' will want to always hire θ', a contradiction. The only exception is if both firms are indifferent about hiring both θ and θ', and this can only happen if actions are constant and equal to $\hat{a}^n$ on the tied interval, and profits are dissipated.

Say that s has *strictly positive sorting (SPS)* if $\theta_h^n = \theta_l^{n+1}$ for all $n \in \{1, \ldots, N-1\}$, so that there are no regions of ties. Under SPS, there will often be gaps in the effort level induced (or set of products offered) as one moves from one firm to the next. Indeed, a gap seems the generic outcome, as v^n must cross v^{n+1} from below at θ_h^n, and only for carefully chosen parameters will this crossing be tangential. Figure 1 shows a typical example with SPS and three firms.

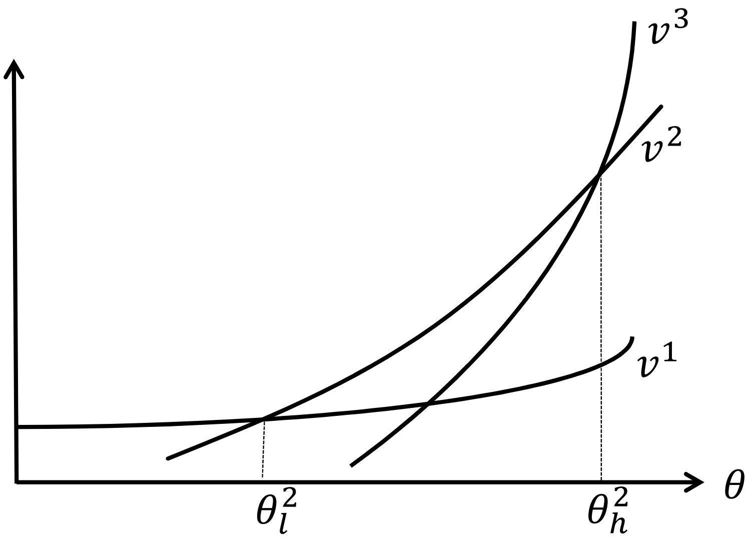

Figure 1: **An example with three firms and strict positive sorting.** Firm 2 hires workers between θ_l^2 and θ_h^2, with Firm 1 hiring lower workers, and Firm 3 higher workers.

4.3.1 No Extraneous Offers

Without some refinement, one can have equilibria in which firm n wins almost always on (θ_l^n, θ_h^n), but at some zero measure set of points in (θ_l^n, θ_h^n), $\varphi^n(\theta) = 1/2$, since $v^n(\theta)$ is equal, for example, to $v^{n+1}(\theta)$. While these offers by $n+1$ may lose money when accepted, they do not hurt Firm $n+1$, since they have measure zero. Similarly, there may be discontinuities in α^{n+1} outside of the region where $n+1$ wins that create complicated incentives for other firms.

For technical and aesthetic reasons, we wish to rule out these difficulties. In Lemma 5 we show that any best response for n must be continuous on $\left(\theta_h^{n-1}, \theta_l^{n+1}\right)$, the region over which n ever wins. We will consider equilibria in which each α^n is continuous everywhere, and in which $\alpha^{n+1} - \alpha^n > 0$ outside of $[\theta_h^{n-1}, \theta_l^{n+1}]$, a condition we will term *no extraneous offers* (NEO). Condition NEO will hold, for example, if each n offers the same action to types above θ_l^{n+1} that it offers to θ_l^{n+1} and the same action to types below θ_h^{n-1}that it offers to θ_h^{n-1}.[12]

4.4 Internal Optimality (IO)

Consider the situation of Firm 2 in Figure 1. Assume first that v^2 reflects a situation where the effort level is everywhere efficient. How can Firm 2 improve its profits? One way might be to change the boundaries of the interval $I = \left[\theta_l^2, \theta_h^2\right]$ of types hired, a topic we will turn to in the next section. But, even holding fixed I, Firm 2 would like to reduce the information rents of types

[12]It is an open question whether there are interesting settings in which this refinement rules out existence.

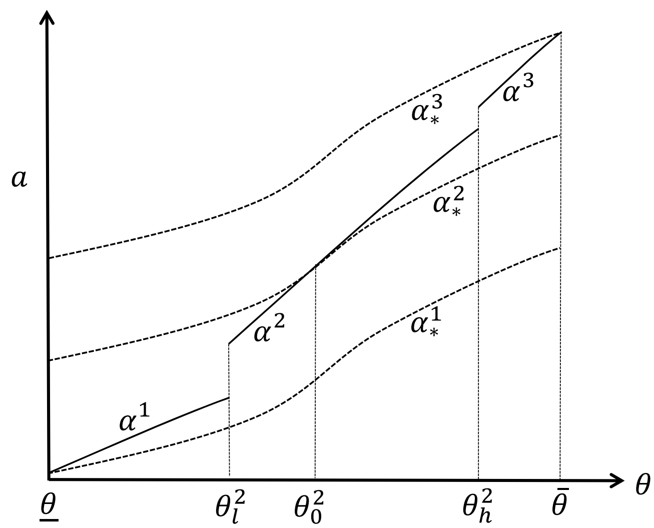

Figure 2: **Efficient and Equilibrium Effort Schedules.** Effort is distorted upward for Firm 1, downward for Firm 3, and first downward and then upward for Firm 2.

in the middle of I. It can do this by making v^2 flatter for some range of types on the bottom "half" of I, while making v^2 steeper on the top "half." This makes v^2 more convex, pushing it downward in the middle of I. Note that in so doing, the firm is distorting effort downward from the efficient level on the lower part of I, but it is distorting effort *upwards* on the upper part of I.

Consider instead Firm 3. As drawn, Firm 3 faces a binding utility constraint only on the lowest type it hires. To lower the utility of other types, it lowers the slope of v^3, which it accomplishes by distorting effort downward for all types except the highest, just as a monopolist would in the standard screening problem (Mussa and Rosen (1978),Maskin and Riley (1984)). For the lowest firm, the situation is the reverse–the utility constraint that binds is that of the *highest* type hired. The way to lower utility on types below this type is to make v^1 *steeper*. That is, Firm 1 distorts actions *upwards* on all types except its lowest.

In Figure 2 we illustrate these distortions in effort. The dotted curves represent the efficient effort schedules, α_*^n, for the three firms, where given our assumptions, efficient effort is higher (at each type) for higher firms. The solid line labeled α^1 is the portion of Firm 1's equilibrium effort schedule that is accepted in equilibrium. It begins at the efficient effort level $\alpha_*^1(\underline{\theta})$ and then lies strictly above α_*^1, reflecting the upward distortion of effort (or quality). Similarly, Firm 3's equilibrium effort schedule is distorted everywhere downward. Finally, α^2 single-crosses α_*^2 from below at θ_0^2, so that equilibrium effort is distorted downward for types below θ_0^2 and upwards for types above. In this example, no firm asks for an effort level between $\alpha^1(\theta_l^2)$ and $\alpha^2(\theta_l^2)$, or between $\alpha^2(\theta_h^2)$ and $\alpha^3(\theta_h^2)$. These gaps are typical in the case of SPS.

To see the pattern of effort distortions, we need a definition. Fix n, and define γ^n by

$$\pi_a^n(\theta, \gamma^n(\theta, \kappa), v) = \frac{\kappa - H(\theta)}{h(\theta)}. \tag{3}$$

Note that γ^n is strictly decreasing in κ. Strategy profile s satisfies *internally optimality* (IO) if for each n, there is $\kappa^n \in [(H(\theta_l^n), H(\theta_h^n)]$, where $\kappa^1 = 0$ and $\kappa^N = 1$.[13]

For expositional convenience, we assume that (3) has an interior solution. One way to do ensure this is to assume that $\lim_{a\to 0} B_a^n(a) = \infty$, and $\lim_{a\to\infty} B_a^n(a) = -\infty$. Another is to assume that $\lim_{a\to 0} B_a^n(a) = \infty$, $\lim_{a\to\infty} B_a^n(a) = 0$, and $\left(1-\bar{\theta}\right) h\left(\bar{\theta}\right) > 1$.[14]

To see at intuitive level that IO gives the right pattern of distortions note first that for firm N, (3) reduces to the standard equation (Mussa and Rosen (1978),Maskin and Riley (1984)) for a monopolist screening an agent of unknown type, an intuition we will generalize below.

Consider an interior firm n. We have argued intuitively above that the firm will first distort downwards and then upwards. Let θ_0 be the dividing point between these regions, so that at θ_0, the action is efficient. Now, consider any $\theta \in [\theta_l, \theta_0]$. Ignore for a moment the monotonicity constraint on actions, and consider raising effort a little at θ, but simultaneously lowering it at θ_0 by the same amount. This leaves the utility at θ_l and θ_h unaffected, and so the same types are hired as before. Changing the effort level of type θ directly changes the payoff to the firm at rate $\pi_a(\theta, \alpha, v)\, h(\theta)$. Since the effort of type θ_0 was efficient, changing type θ_0's effort a little has no direct impact on the firm's payoff. Finally, the utility of all types between θ and θ_0 is raised at rate $-c_{a\theta} = 1$, for cost $H(\theta_0) - H(\theta)$. At an optimal profile, this perturbation must have zero impact, and so, setting benefit equal to cost and dividing by $h(\theta)$, we have that the optimal action at θ must satisfy

$$\pi_a(\theta, \alpha, v) = \frac{H(\theta_0) - H(\theta)}{h(\theta)},$$

which is (3) with $\kappa = H(\theta_0)$. The argument when $\theta > \theta_0$ is very similar.

Let us now see how to do this at a more formal level. Fix boundary points θ_l and θ_h, and let $\mathcal{P}(\theta_l, \theta_h)$ be the optimization problem given by

[13] By Lemma 6 in Section 8.4, log-concavity of H and $1-H$ imply that $(\kappa - H(\cdot))/h(\cdot)$ is decreasing for all $\kappa \in [0,1]$, and so, since $\pi_{a\theta} = 1 > 0$, $\gamma^n(\cdot, \kappa)$ is indeed strictly increasing under IO.

[14] Since $c_a = 1-\theta$, we have $B_a(\gamma^n(\theta, \kappa)) = ((\kappa - H(\theta))/h(\theta)) + 1 - \theta$ which is minimized at $\theta = \bar{\theta}$ and $\kappa = 0$, which is positive if $\left(1-\bar{\theta}\right) h\left(\bar{\theta}\right) > 1$. Hence, since B_a has range $(0,\infty)$, (3) has an interior solution.

$$\max_{(\alpha,v)} \quad \int_{\theta_l}^{\theta_h} \pi(\theta,\alpha,v)h(\theta)d\theta \tag{4}$$

$$s.t. \quad v(\theta_l) \geq v^{-n}(\theta_l) \tag{5}$$

$$v(\theta_h) \geq v^{-n}(\theta_h), \text{ and} \tag{6}$$

$$v(\theta) = v(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \alpha(\tau)\, d\tau \text{ for all } \theta. \tag{7}$$

This is a relaxation of the problem actually faced by the firm, since we drop monotonicity of α, ignore the value of v except at θ_l and θ_h, and check relaxed versions of the constraint at θ_l and θ_h. We will deal with the optimal choice of θ_l and θ_h in the next section. Note that in $\mathcal{P}(\theta_l,\theta_h)$, the objective function is concave (since $\pi(\theta,\cdot,\cdot)$ is concave), and the set of feasible (α,v) is convex.

Define

$$z(\theta_l,\theta_h,\kappa) = v^{-n}(\theta_h) - v^{-n}(\theta_l) - \int_{\theta_l}^{\theta_h} \gamma(\theta,\kappa)d\theta.$$

That is, $z(\theta_l,\theta_h,\kappa)$ is the difference between the "rise" in v^{-n} between θ_l and θ_h and the rise in v implied by $\gamma(\cdot,\kappa)$ on the same interval. Note that $z(\theta_l,\theta_h,\cdot)$ is increasing, because γ falls in κ.

As suggested by the intuition above, the solution will be of the γ form. We will use z to define the optimal κ (z will also be very useful later in the paper). If both (5) and (6) bind, then κ is tied down by $z(\theta_l,\theta_h,\kappa)=0$. Imagine that $z(\theta_l,\theta_h,H(\theta_h))<0$. It will turn out that in this case, it is optimal in the relaxed problem to set κ to equal $H(\theta_h)$, with (5) binding, and (6) slack. This is in fact what occurs for Firm N, for whom $H(\theta_h)=1$, and for whom only (5) binds. Similarly, if $z(\theta_l,\theta_h,H(\theta_l))>0$, then it will turn out to be optimal to set κ equal to $H(\theta_l)$, with (5) slack and (6) binding, as indeed occurs for Firm 1, for whom $H(\theta_l)=0$.

Given this intuition, define $\tilde{\kappa}(\theta_l,\theta_h)$ as the element of $[H(\theta_l),H(\theta_h)]$ that makes $z(\theta_l,\theta_h,\cdot)$ as close as possible to zero. That is, define $\tilde{\kappa}(\theta_l,\theta_h)$ as $H(\theta_l)$ if $z(\theta_l,\theta_h,H(\theta_l))>0$, as $H(\theta_h)$ if $z(\theta_l,\theta_h,H(\theta_h))<0$, and as the solution to $z(\theta_l,\theta_h,\kappa)=0$ otherwise.

We then have the following solution to the relaxed problem.

Lemma 1 (Relaxed Problem) *Problem $\mathcal{P}(\theta_l,\theta_h)$ has a solution $\tilde{s}(\theta_l,\theta_h)=(\tilde{\alpha},\tilde{v})$. On $[\theta_l,\theta_h]$, $(\tilde{\alpha},\tilde{v})$ is unique and has $\tilde{\alpha}=\gamma(\cdot,\tilde{\kappa}(\theta_l,\theta_h))$.*[15] *If $\tilde{\kappa}(\theta_l,\theta_h)>H(\theta_l)$ then $\tilde{v}(\theta_l)=v^{-n}(\theta_l)$, and if $\tilde{\kappa}(\theta_l,\theta_h)<H(\theta_h)$ then $\tilde{v}(\theta_h)=v^{-n}(\theta_h)$.*

To see the intuition, start from the case where (6) is slack. Raise effort a little at any given θ in (θ_l,θ_h). This gains $\pi_a(\theta,\alpha,v)$ on $h(\theta)$ workers, but raises the surplus of the $H(\theta_h)-H(\theta)$ workers above θ. For this not to be profitable, we must have $\pi_a(\theta,\alpha,v)h(\theta)-(H(\theta_h)-H(\theta))=0$, or equivalently, $\tilde{\alpha}=\gamma(\cdot,H(\theta_h))$. Similarly, if (5) is slack, then $\kappa=H(\theta_l)$.

[15] Since we have imposed continuity on action schedules, there is no difference between two action schedules agreeing "everywhere" and "almost everywhere."

It cannot be that both (5) and (6) are slack, since then a reduction of v by a constant is profitable. So, we are left with the case where both (5) and (6) bind at the optimum. Then, as argued above, the solution must be of the form given by (3), where κ is tied down uniquely by (5) and (6). Note that it cannot be that $\kappa > H(\theta_h)$, since then it is better to lower κ to $H(\theta_h)$, making (6) slack, and similarly it cannot be that $\kappa < H(\theta_l)$ since then it is better to raise κ to $H(\theta_l)$, making (5) slack. In particular, these are the solutions to the further relaxed problem where one of the constraints need not be an equality.

To prove that optimal solution in the original problem is also of this form, assume that (α, v) does not agree with $(\tilde{\alpha}, \tilde{v})$ on (θ_l, θ_h). We begin by perturbing (α, v) linearly in the direction of $(\tilde{\alpha}, \tilde{v})$, but modify the perturbation to keep payoffs greater than v^{-n} on (θ_l, θ_h), so that the firm continues to hire those workers. Doing this while maintaining monotonicity of α may mean hiring some workers outside of (θ_l, θ_h). Using PP, we next purge any unprofitable workers. We show that with these two modifications, the initial impact of this perturbation is at least as profitable as simply moving in the direction of $(\tilde{\alpha}, \tilde{v})$. But, since $(\tilde{\alpha}, \tilde{v})$ is the unique solution to $\mathcal{P}(\theta_l, \theta_h)$, and since the objective function in $\mathcal{P}(\theta_l, \theta_h)$ is concave, moving in this direction raises profits, a contradiction. We will show in the next section that for $n \notin \{1, N\}$, κ is interior in equilibrium.

An economic implication of IO is that there is complete sorting of workers–or, alternatively, a complete product line–within the interval of types served by each firm. This of course depends on the absence of a fixed cost per menu item.

4.5 Optimal Boundaries (OB)

Strategy profile s satisfies the *optimal boundary condition* (OB) if

$$\pi^n(\theta_l^n, \alpha^n, v^n) - \pi_a^n(\theta_l^n, \alpha^n, v^n)(\alpha^n(\theta_l^n) - a^{-n}(\theta_l^n)) = 0, \text{ and} \tag{8}$$

$$\pi^n(\theta_h^n, \alpha^n, v^n) + \pi_a^n(\theta_h^n, \alpha^n, v^n)(a^{-n}(\theta_h^n) - \alpha^n(\theta_h^n)) = 0, \tag{9}$$

where (8) is discarded for Firm 1, and (9) for Firm N.

Each equation balances the direct profit on the boundary worker and a term which is the product of (i) the marginal profit of requiring a higher action of the boundary worker and (ii) the difference in action between firm n and the adjacent firm at the boundary type. By PS and NEO, these differences are positive. In contrast to NP, which is about stealing potentially distant workers, OB reflects that small changes in the set of workers employed must not be profitable.

To see the intuition for OB, fix n and increase the effort of types near θ_h a little. This has direct benefit $\pi_a(\theta_h, \alpha, v)h(\theta_h)$. But, as $v(\theta_h)$ is raised, θ_h is increased at rate $1/(a^{-n}(\theta_h) - \alpha(\theta_h))$ and so the profit on the new workers hired is $\pi(\theta_h, \alpha, v)h(\theta_h)/(a^{-n}(\theta_h) - \alpha(\theta_h))$. Cancelling $h(\theta_h)$ and rearranging yields (9). The derivation of (8) is similar, noting that to lower θ_l and gain extra workers, one *reduces* effort on types near θ_l, holding fixed the surplus of higher types.

The definition of OB discards (8) for $n = 1$ and (9) for $n = N$, rather than replacing them with inequalities. The reason is that given the above discussion, IO implies that, holding fixed θ_h, Firm 1 is better off with $\kappa = 1$ and $\theta = \underline{\theta}$ than with any higher θ_l. Hence, checking optimality of θ_h for Firm 1 is enough, and similarly it is enough to check the optimality of θ_l for N.

We will use our next simple lemma repeatedly. The slope of profit with respect to θ has the sign of $\pi_a \alpha_\theta$, and if the action profile is of the γ form, then profits are strictly single-peaked.

Lemma 2 *For any* $(\alpha, v) \in S^n$*,*

$$(\pi(\theta, \alpha, v))_\theta = \pi_a(\theta, \alpha, v)\, \alpha_\theta(\theta), \tag{10}$$

and if $\alpha = \gamma(\cdot, H(\theta_0))$ *on* $[\theta_l, \theta_h]$*, with* $\theta_0 \in [\theta_l, \theta_h]$ *then* $\pi(\cdot, \alpha, v)$ *is strictly single-peaked on* $[\theta_l, \theta_h]$*, with peak at* θ_0*.*

To see the proof of (10) note that by (1)

$$\pi_\theta(\theta, \alpha, v) = -c_\theta(\alpha(\theta), \theta) = \alpha(\theta) = v_\theta(\theta) = -\pi_v(\theta, \alpha, v)\, v_\theta(\theta),$$

and so only the effect through a remains. If $\alpha = \gamma(\cdot, H(\theta_0))$ on $[\theta_l, \theta_h]$, then from (3) first $\alpha_\theta > 0$, and second $\pi_a(\theta, \alpha, v)$ has strictly the same sign as $\theta_0 - \theta$. Hence, π is strictly single-peaked at θ_0.

With OB and Lemma 2 in hand, let us see that κ is interior for $n \notin \{1, N\}$. Assume $\kappa = H(\theta_h)$. Then by Lemma 2, $\pi(\cdot, \alpha, v)$ is strictly increasing on (θ_l, θ_h). But, $\pi(\theta_l, \alpha, v) \geq 0$ and hence $\pi(\theta_h, \alpha, v) > 0$. But, since $\kappa = H(\theta_h)$, we also have $\pi_a(\theta_h, \alpha, v) = 0$, and so (9) is violated. Essentially, if $\kappa = H(\theta_h)$, then increasing effort on types near θ_h has second-order efficiency costs but gains some extra workers on whom profits are strictly positive. Similarly, $\kappa > H(\theta_l)$.

Recall that θ_0, the point at which $H(\theta_0) = \kappa$, is equal to $\underline{\theta}$ for $n = 1$, is in (θ_l, θ_h) for $n \in \{2, \ldots, N-1\}$, and is equal to $\bar{\theta}$ for $n = N$. The fact that profits are strictly single-peaked at θ_0 has some intuition: For intermediate firms, customers in the middle of the participation range find neither of the alternative firms very attractive, and so are the easiest to extract rents from. Similarly, for the end firms, it is the extreme types from whom it is easiest to extract rents.

One key implication of Lemma 2 is that in equilibrium, profits π are strictly positive everywhere on the region (θ_l, θ_h) where the firm is uniquely active. This follows since by PP, π is positive at θ_l and θ_h, and since α is of the γ form on $[\theta_l, \theta_h]$, and so π is strictly single-peaked on $[\theta_l, \theta_h]$.

How about profits on the boundary types θ_l and θ_h? We have argued that if there is a region of overlap between the two firms, then profits on these types are driven to zero. Consider the case depicted in Figure 1, where the surplus functions cross strictly, and so the implemented action jumps at the boundary. As argued, this is the normal case under SPS, and is always the case when firms are sufficiently differentiated (see Section 6.1). But then, since we have already argued that $\theta_0 > \theta_l$ for $n \geq 2$, the term $\pi_a(\theta_l, \alpha, v)(\alpha(\theta_l) - a^{-n}(\theta_l))$ in (8) will be strictly positive. Thus,

$\pi(\theta_l, \alpha, v)$ must be strictly positive, and similarly for $\pi(\theta_h, \alpha, v)$. Even though firms compete for the boundary customer, the Bertrand logic does not imply zero profits, since the difference in their production technologies implies that neither firm can profitably imitate the other.

5 Other Implications of Necessity

5.1 Discounts and (Non-)Implementation by Linear Contracts

Fix n, and let the tariff T paid by the *firm to the worker* associated with action a be implicitly defined by $T(\alpha(\theta)) = v(\theta) + c(\alpha(\theta), \theta)$. Then,

$$T_a(\alpha(\theta))\alpha_\theta(\theta) = v_\theta(\theta) + c_a(\alpha(\theta), \theta)\alpha_\theta(\theta) - \alpha(\theta) = c_a(\alpha(\theta), \theta)\alpha_\theta(\theta) = (1-\theta)\,\alpha_\theta(\theta)$$

and hence, $T_a(\alpha(\theta)) = 1 - \theta$. But then, $T_{aa}(\alpha(\theta)) = -1/\alpha_\theta(\theta)$, and so T is strictly concave in a. It follows first that there are 'quantity discounts': the wage per unit of effort decreases in the amount of effort, and hence higher types obtain a lower wage per unit of effort. Further, since T is strictly concave, it cannot be implemented using a menu of its tangents, that is, using linear contracts (see Laffont and Martimort (2002), Section 9.5).

Similarly, in the product interpretation of the model the amount paid by the *consumer to the firm* is $\tilde{T}(\alpha(\theta)) = -v(\theta) - c(\alpha(\theta), \theta)$, which, arguing as before, is strictly convex, and hence once again cannot be implemented by a menu of linear contracts. It can also be shown that $\tilde{T}/a$ increases in a, and therefore there are quantity premia.

5.2 The Competitive Limit

We now explore the behavior of our economy as N grows. Let $d_1 = \max_{a,n}(B^n(a) - \max_{n' \neq n} B^{n'}(a))$. When d_1 is small, then for any firm n and action a, there is another firm for whom $B^{n'}(a)$ is nearly as large as $B^n(a)$. Also, for each firm n, define (a_l^n, a_h^n) as the interval of actions over which firm n is the most efficient, i.e., over which $B^n(a) > \max_{n' \neq n} B^{n'}(a)$, and define $d_2 = \max_n(a_h^n - a_l^n)$ as the longest such interval. Each of d_1 and d_2 is a measure of how far apart the firms are.

Example 1 *Let $B(a, \tau) = a - (a - \tau)^2$ and for each $n \in \{1, \ldots, N\}$ let $B^n(a) = B(a, n/N)$. Then, $d_1 = 1/N^2$ and $d_2 = 1/N$.*

In this example, as N grows large, d_1 and d_2 both converge to 0, and do so quickly. In general, d_1 and d_2 will be small in economies with many firms that are "spread out," and will converge to zero quickly if they are spread "evenly."[16]

[16] For an example where convergence fails, start from a two-firm example, and then create $N-1$ copies of Firm 1 while retaining a single copy of Firm 2. For an example with slow convergence spread $\lfloor\sqrt{N}\rfloor$ firms out evenly as in Example 1, and make the remainder copies of Firm 1.

Our main result of this section is that as $d_1 \to 0$ and $d_2 \to 0$, the payoff to both firms and workers converges to the competitive limit. Let $\delta = \max_{a,n,\theta}(c_{aa}(a,\theta) - B^n_{aa}(a))$ bound the absolute value of $B^n_{aa} - c_{aa}$.

Theorem 2 (Limit Efficiency) *Let s be an equilibrium. Fix θ, and let n be the firm that serves θ. Then*

$$0 \leq \pi^n(\theta, \alpha^n, v^n) \leq B^n(\alpha^n(\theta)) - \max_{n'} B^{n'}(\alpha^n(\theta)) \leq d_1, \tag{11}$$

and

$$v_*(\theta) - v^n(\theta) \leq d_1 + \frac{1}{2} d_2^2 \delta. \tag{12}$$

In Example 1, $\pi^n(\theta, \alpha^n, v^n) \leq 1/N^2$, and $v_*(\theta) - v^n(\theta) \leq (1 + \frac{1}{2}\delta)/N^2$, and so convergence is fast.

The first inequality in (11) follows directly from PP. The second inequality follows from NP: if Firm n is earning above this bound, then by definition of d_1, there is some other firm that can profitably imitate them. By (11), conditional on the effort level asked of θ, the match between the firm and the agent is efficient. Note, however, that the firm to whom one is matched in equilibrium may not be the firm that is optimal conditional on *efficient* effort. That is, $\arg\max_{n'} B^{n'}(\alpha_*^{n'}(\theta)) - c(\alpha_*^{n'}(\theta), \theta)$ need not equal n. There are thus three sources that pull the surplus of the agent down from the competitive equilibrium level. First, effort will typically be *distorted* from $\alpha_*^n(\theta)$. Second, the worker may be *mismatched*. Third, the firm to whom the worker is matched earns *rents*.

To see the intuition for (12), note that if $\hat{n}$ is the firm that serves θ efficiently, then θ can imitate some type $\hat{\theta}$ that $\hat{n}$ does serve in equilibrium, say at effort level $\hat{a}$, and, using (11), earn within d_1 of the surplus generated by that match. But, using (11), $\hat{a}$ and $\alpha_*^{\hat{n}}(\theta)$ must both be actions where $\hat{n}$ is the most efficient firm. Thus, $\hat{a}$ and $\alpha_*^{\hat{n}}(\theta)$ are at most d_2 apart, and so the difference in the match surplus generated by these actions is correspondingly small, indeed of order d_2^2, since $\alpha_*^{\hat{n}}(\theta)$ maximizes match surplus.

5.3 Who Does Asymmetric Information Help or Hurt?

Consider the version of our model where one removes the workers' private information. Under monopoly the effect of this removal is clear: the firm is better off, since it can undo any inefficiency, raising total surplus, and then extract all the surplus as information rents disappear. The workers, who now earn no rents, are clearly worse off.

In the oligopoly case, there is a third effect. With asymmetric information, worker θ might be hired by firm n, even though some other firm n' could, by an appropriate choice of wage and effort, both attract and earn profits on θ. What holds n' back is that such an offer might also attract some of n''s existing workers at lower profits. Without asymmetric information, there is no such cross-type constraint, making it easier to compete for a worker currently hired by a competitor. Hence, while θ no longer receives information rents, his *outside option* may have increased.

To examine this issue, note first that equilibrium with perfect information in the oligopoly case suffers from the classic discontinuity-at-ties problem. To sidestep this, assume that when the worker faces two offers giving him the same surplus, he chooses the firm that earns more surplus in hiring him. Also, to avoid unnatural equilibria, assume that no firm makes an offer that they would lose money on if accepted. Then, competition at each type is Bertand between differentiated firms, and so in equilibrium each worker will be hired by the firm that can use him best, effort will be efficient, and surplus will equal the surplus that the second most efficient firm for that type can provide. Since the allocation is efficient, positive sorting holds.

Theorem 3 (Welfare) *Let $\hat{\theta}$ be on the boundary between the regions of types efficiently hired by two consecutive firms. Then, for all θ in some interval containing $\hat{\theta}$, θ is strictly worse off, and the firm hiring θ strictly better off, under asymmetric information than under full information.*

The proof is simple. Since $\hat{\theta}$ can be efficiently hired by two firms, the Bertrand logic implies that in the full-information equilibrium, $\hat{\theta}$ earns the efficient surplus, and the firms earn zero. In the asymmetric information case, we have already proven that firms earn strictly positive profits, and so, since total surplus is at most the efficient surplus, the surplus of the worker must be strictly lower than in the full-information equilibrium. The argument is completed by noting that profits and surplus are continuous in type.

That is, *contrary* to the case of a monopoly firm, it is workers who are harmed by asymmetric information, and firms who are helped, at least over ranges of types near points where two firms can offer the efficient surplus. We do not have clear results or intuition for how the two forces–one in favor of more competition and the other against–balance outside of these ranges.

6 Existence and Sufficiency

To complete the analysis, we now turn to sufficiency and existence. We will provide a set of conditions for a strategy profile s to be an equilibrium, and for an equilibrium to exist.

6.1 Stacking and Strict Regularity

The possibility of ties at the boundaries between players substantially complicates things. So, we begin by imposing some simplifying structure on the problem.

Definition 1 Stacking *is satisfied if for all $n < N$, $\gamma^{n+1}(\cdot, 1) > \gamma^{n}(\cdot, 0)$.*

Under stacking, when each κ is restricted to lie in $[0, 1]$, the action schedule for player $n + 1$ always lies strictly above that of player n. Stacking holds if firms are sufficiently differentiated.[17] Stacking simplifies our analysis since in any strategy profile that will be relevant to us, the surplus functions of adjacent players will cross strictly, precluding ties. Furthermore, because the crossings are strict, a small change in the strategy of one player will change the set of types hired in a continuous fashion, getting rid of a key discontinuity.[18] *We will henceforth impose stacking.*

6.2 The Main Results

We will state the main results first, and then, in the next several subsections, discuss how to prove them. The relevant proofs are in Appendix B.

Fix n and s^{-n}. We first need a definition of what it means for two strategies for n to differ only in inessential ways given s^{-n}. Say that $s^n = (\alpha^n, v^n)$ and $\hat{s}^n = (\hat{\alpha}^n, \hat{v}^n)$ are *essentially equivalent* if $\varphi(\cdot, (s^n, s^{-n})) = \varphi(\cdot, (\hat{s}^n, s^{-n}))$, and if anywhere that $\varphi(\cdot, (s^n, s^{-n})) > 0$, we have $\alpha^n = \hat{\alpha}^n$ and $v^n = \hat{v}^n$. That is, s^n and $\hat{s}^n$ agree anywhere that is relevant given s^{-n}. Two strategy profiles are essentially equivalent if they are essentially equivalent for each n.

Theorem 4 (Sufficiency) *Assume stacking. Then any strategy profile satisfying PS, IO, and OB is essentially equivalent to a Nash equilibrium.*

This is non-trivial, because $\Pi^n(\cdot, s^{-n})$ is not quasi-concave: if we fix s^{-n}, s^n, and $\hat{s}^n$, then, a convex combination of s^n and $\hat{s}^n$ will typically win a set of types different from either s^n or $\hat{s}^n$, and so it unclear how its profits will relate to those of either s^n or $\hat{s}^n$.[19] But then satisfying the first-order conditions need not imply optimality.

Theorem 5 (Existence) *Assume stacking. Then, a Nash equilibrium exists.*

Existence is not trivial since Π is not continuous on S. For example, let $N = 2$ and $v^2 = v^1 + \varepsilon$. Then $\varphi^2(\cdot, s) = 0$ for all $\varepsilon < 0$, while $\varphi^2(\cdot, s) = 1$ for all $\varepsilon > 0$. Further, since Π is not quasi-concave, the set of best-responses may be non-convex.

6.3 The Reformulation

Let us reformulate the problem of finding a best response. Fix n and s^{-n}. Strategy s^n is *dominant* on (τ_l, τ_h) if (τ_l, τ_h) is a maximal interval such that $v^n > v^{-n}$. Say that s^n is *single dominant*

[17] For example, if θ is uniform on $[0, 1/3]$, and if $B^n(a) = \zeta^n + \beta^n \log a$, then it can be verified that stacking will hold as long as $\beta^{n+1}/\beta^n > 2$.

[18] If firms are not very differentiated, then equilibria must involve intervals of ties. To see this, consider $N = 2$, and assume that $\gamma^2(\cdot, 1) < \gamma^1(\cdot, 0)$. Then, there must be an interval over which the firms are tied (with associated action equal to $\hat{a}_1$), since if $\theta^2_l = \theta^1_h$, then $\alpha^1(\theta^1_h) = \gamma^1(\theta^1_h, 0) > \gamma^2(\theta^1_h, 1) = \alpha^2(\theta^1_h)$, contradicting PS.

[19] Note that for a given θ, π is strictly concave in s^n.

on (τ_l, τ_h) if in addition $v^n < v^{-n}$ for $\theta \notin [\tau_l, \tau_h]$. That is, Firm n wins with probability one on (τ_l, τ_h), and probability zero outside of $[\tau_l, \tau_h]$.

The first key step is to show that if the other firms are doing something "reasonable," then the firm can optimize over single-dominant strategies that are of the γ form, with $\kappa \in [0, 1]$. To formalize "reasonable" note first that while the convex combination of two γ strategies each with $\kappa \in [0, 1]$ need not be a γ strategy, it will always satisfy the following condition.

C1 α^n is continuous, with $\alpha^n(\theta) \in [\gamma^n(\theta, 1), \gamma^n(\theta, 0)]$ for all θ.

Given Proposition 3 in the Appendix, it is also innocuous to assume that firms never offer a surplus above v_*^n, the most surplus they can offer without losing money.

C2 $v^n \leq v_*^n$.

By $C2$ and (2) it follows that n, in any best response to s^{-n}, earns strictly positive profits. Consider any s^{-n} that satisfies $C1$. Then, by stacking, all actions by competitors below n are below $\gamma^n(\cdot, 1)$, and all actions by competitors above n are above $\gamma^n(\cdot, 0)$. Hence, there is $\theta^x \in [\underline{\theta}, \bar{\theta}]$ such that $a^{-n} < \gamma^n(\cdot, 1)$ for $\theta < \theta^x$, and $a^{-n} > \gamma^n(\cdot, 0)$ for $\theta > \theta^x$. In Figure 1, and from the perspective of Firm 2, θ^x is the point at which v^1 and v^3 cross.

Lemma 3 *Assume stacking, let s satisfy $C1$, and assume that n sometimes wins. Then, s^n is single dominant on some non-empty interval including θ^x, and if s^n satisfies OB, it satisfies NP as well.*

That s^n is single dominant on some non-empty interval including θ^x follows since by $C1$ and stacking, v^n can only cross v^{-n} twice, once below θ^x and once above, and these crossing are strict.

The second result follows since a^{-n} is above the efficient level for n to the right of θ_h, and hence by (10), the profits to poaching are decreasing. And, we show that near θ_h, (9) implies that n prefers to gain an extra worker by moving θ_h than by poaching. The proof is similar for $\theta < \theta_l$.

Corollary 1 *Under stacking, any equilibrium has SPS.*

This follows immediately since we have already shown that in any equilibrium, all players sometimes win, and that their strategies are of the γ form with $\kappa \in [0, 1]$, and hence satisfy $C1$.

6.4 Relating the Original and the Relaxed Problem

Our goal is to move the analysis of n's problem from the original infinite-dimensional problem of choosing an entire action schedule and associated surplus function to a more tractable two-dimensional problem. Effectively, each firm just optimizes over the endpoints of the region over which it is winning, knowing that for any given endpoints, the action schedule will be of the $\gamma(\cdot, \kappa)$ form with κ and surplus tied down by the endpoint utilities.

To formalize this, recall from Section 4.4 that $\tilde{s}(\theta_l, \theta_h)$ is the solution to the relaxed problem $\mathcal{P}(\theta_l, \theta_h)$, where the action profile is $\gamma(\cdot, \tilde{\kappa}(\theta_l, \theta_h))$, with $\tilde{\kappa}(\theta_l, \theta_h) \in [H(\theta_l), H(\theta_h)]$. Let $r(\theta_l, \theta_h)$ be the resulting value of $\mathcal{P}(\theta_l, \theta_h)$. We wish to relate the maximization of r to the maximization of Π^n, the profits to Firm n in the original problem. This is accomplished in the next three claims.

The first claim establishes that r has a maximum (θ_l, θ_h), and that for any maximum of r, the associated solution to the relaxed problem is feasible in the original game, hires the interval of types (θ_l, θ_h), and has the same payoff as r.

Lemma 4 *Assume stacking. Fix n, and let s^{-n} satisfy $C1$ and $C2$. Then r has a maximum, and at any maximum (θ_l, θ_h) of r,*

(i) $\tilde{s}(\theta_l, \theta_h) \in S$,

(ii) if $\theta_l > \underline{\theta}$, then $\tilde{v}(\theta_l) = v^{-n}(\theta_l)$ and if $\theta_h < \bar{\theta}$, then $\tilde{v}(\theta_h) = v^{-n}(\theta_h)$, and

(iii) $\tilde{s}(\theta_l, \theta_h)$ is single dominant on (θ_l, θ_h) with $\Pi(\tilde{s}(\theta_l, \theta_h), s^{-n}) = r(\theta_l, \theta_h)$.

A maximum exists since r is continuous on a compact set. Part (i) follows since $\tilde{\kappa}(\theta_l, \theta_h) \in [0, 1]$, so that $\gamma(\cdot, \tilde{\kappa})$ is increasing. The key to the proof of (ii) is to show that if for example $\tilde{v}(\theta_h) > v^{-n}(\theta_h)$, then $r_{\theta_h}(\theta_l, \theta_h) > 0$, which, since (θ_l, θ_h) is optimal, implies $\theta_h = \bar{\theta}$. Part (iii) follows immediately from (ii).

The next claim is that for any strategy the firm might contemplate in the original game, there is a pair (θ_l, θ_h) such that $r(\theta_l, \theta_h)$ is at least as big as the payoff to that strategy.

Proposition 2 *Assume stacking. Fix n and s^{-n} satisfying $C1$ and $C2$. Then, for each $\hat{s}$ there is (θ_l, θ_h) with $\Pi(\hat{s}, s^{-n}) \leq r(\theta_l, \theta_h)$.*

Before we discuss the proof, we note that Lemma 4 and Proposition 2 between them justify the desired reparameterization:

Corollary 2 *Assume stacking. Fix n and s^{-n} satisfying $C1$ and $C2$. Then, $\hat{s}$ is a maximum of $\Pi(\cdot, s^{-n})$ if and only if $\hat{s} = \tilde{s}(\theta_l, \theta_h)$, where (θ_l, θ_h) maximizes r.*

The corollary follows since by Proposition 2, for any strategy $\hat{s}$ Firm n is considering, there is (θ_l, θ_h) for which $r(\theta_l, \theta_h)$ is as big as the payoff to $\hat{s}$. But, by Lemma 4, the strategy associated with $r(\theta_l, \theta_h)$ is feasible and generates payoff $r(\theta_l, \theta_h)$

The proof of Proposition 2 uses two lemmas. Lemma 8 shows that there is an interval $[\underline{m}, \overline{m}]$ of types such that n makes money imitating his opponent if and only if $\theta \in [\underline{m}, \overline{m}]$, and where $\theta^x \in [\underline{m}, \overline{m}]$. This follows from (10) since by $C1$ and stacking, a^{-n} is first strictly below n's efficient action level and then strictly above, and so profits to imitation are single-peaked at θ^x.

Using Proposition 3, assume that $\hat{s} = (\alpha, v)$ never loses money, and is dominant on some interval (τ_l, τ_h). Then Lemma 9 shows that (τ_l, τ_h) and $[\underline{m}, \overline{m}]$ overlap. For intuition, assume that $\tau_l \geq \bar{m}$. Then, we show that since the firm loses money with a^{-n} and v^{-n}, it $a\ fortiori$ loses money with menu items that implement an even more inefficiently high action and offer even more surplus. To see this in more detail, note that by the definition of dominance $v(\tau_l) = v^{-n}(\tau_l)$, but $v(\tau) > v^{-n}(\tau)$ just to the right of τ_l, and so, for some τ' just to the right of τ_l, v is steeper than v^{-n}, and hence $\alpha(\tau') > a^{-n}(\tau')$. Since $\bar{m} \geq \theta^x$, we have that $a^{-n}(\tau')$ is already inefficiently high for n, and so since $\alpha(\tau') > a^{-n}(\tau')$, Firm n must be losing money at τ', a contradiction to the assumption that $\hat{s}$ never loses money.

Armed with these two lemmas, let us see that any $\hat{s}$ is dominated by some strategy that is single-dominant, proving Proposition 2. To do so, let $\overline{m}^* \geq \overline{m}$ capture any region of dominance of v that contains $\overline{m}$, and let $\underline{m}^* \leq \underline{m}$ similarly capture any region of dominance of v that contains $\underline{m}$. Relative to $\hat{s}$, the firm strictly benefits by removing any worker outside of $[\underline{m}^*, \overline{m}^*]$, and by adding any worker in $(\underline{m}, \overline{m})$ that it does not already hire with probability one, since the profits from imitation are strictly positive. But, $\tilde{s}(\underline{m}^*, \overline{m}^*)$ accomplishes exactly this, and does so optimally in the relaxed problem, and hence its associated payoff $r(\underline{m}^*, \overline{m}^*)$ is at least as high as $\Pi(s^n, s^{-n})$.

6.5 Unique Best Responses

In this section, we discuss the building-blocks we will use to prove sufficiency and existence. We will begin by showing that r is sufficiently well-behaved that it has a unique maximum for any given s^{-n} satisfying $C1$ and $C2$, and that any critical point of r is that maximum.

We face three challenges. First, v^{-n} has a kink point at each θ where the relevant opponent changes, and hence so does r. Second, r can have troughs and so single-peakedness fails, complicating a proof of uniqueness. Finally, because our choice set is two dimensional, it is not obvious that single-peakedness alone is enough (recall footnote 3)

Recall that at θ^x, a^{-n} transitions from being driven by opponents with index below n to opponents above n. We begin by showing any optimum of r is in the rectangle $R = [\underline{\theta}, \theta^x] \times [\theta^x, \bar{\theta}]$ illustrated in Figure 3. The proof is in Lemma 13, and relies heavily on Lemma 4.

Note that by $C1$, each kink point of v^{-n} is a point at which one transitions from one opponent to the next, and hence there are at most $N-1$ such points. In Figure 3, $K = \{k_1, k_2\}$. Let $\tilde{R} = [\iota_l, \iota_h] \times [\iota_l', \iota_h']$ be a maximal rectangle with the property that the opponent on (ι_l, ι_h) is constant, the opponent on (ι_l', ι_h') is constant, and $\iota_h \leq \iota_l'$. Using $C1$, v^{-n} is continuously differentiable on $\tilde{R}$, with kinks in r constrained to the boundaries between rectangles.

Recall that $z(\theta_l, \theta_h, \kappa) = v^{-n}(\theta_h) - v^{-n}(\theta_l) - \int_{\theta_l}^{\theta_h} \gamma(\tau, \kappa)\, d\tau$, where since $\gamma_\kappa < 0$, we have $z_\kappa > 0$. Note also that on R,

$$z_{\theta_l} = -a^{-n}(\theta_l) + \gamma(\theta_l, \kappa) > 0, \tag{13}$$

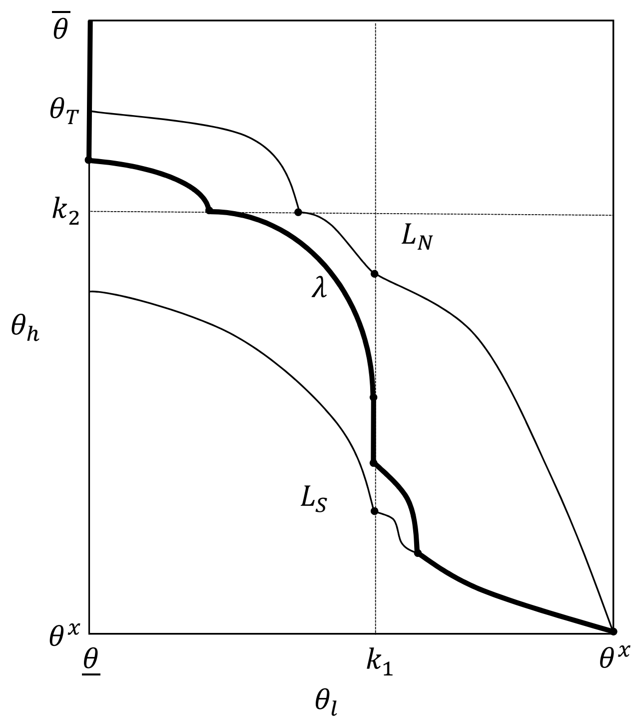

Figure 3: **The rectangle R**. The area between L_S and L_N is Θ. There are kink points in v^{-n} at k_1, k_2, and θ^x. On the four areas delineated by the dotted lines, v^{-n} is continuously differentiable. The thick line is the path described by λ. Where the path runs along L_S, we have $r_{\theta_l} \leq 0$ and $r_{\theta_h} > 0$, and so ψ is increasing. The path never runs along L_N, where $r_{\theta_l} < 0$.

since $\theta_l \leq \theta^x$, and so by $C1$ and stacking, $a^{-n}(\theta_l) < \gamma(\theta_l, \kappa)$. Similarly, $z_{\theta_h} > 0$.

Let the locus L_N be defined by $z(\theta_l, \theta_h, H(\theta_l)) = 0$, and the locus L_S be defined by $z(\theta_l, \theta_h, H(\theta_h)) = 0$. These are the north and south boundaries of the set

$$\Theta = \{(\theta_l, \theta_h) \in R | z(\theta_l, \theta_h, \tilde{\kappa}(\theta_l, \theta_h)) = 0\}.$$

Assume first that L_S hits the western boundary of R, let $\theta_T \leq \bar{\theta}$ be the latitude at which L_N hits the boundary of R, and let A be the (possibly empty) segment of the western boundary of R above θ_T. Using Proposition 1 and 4, we show that any maximum of r occurs either in Θ, with both the utility constraints (5) and (6) binding, or in A, with (5), at $\underline{\theta}$, slack.

Next we show (Lemma 12) that, on any given $\tilde{R} \cap \Theta$, if $r_{\theta_l} = 0$ then r is locally strictly concave in θ_l. Similarly, if $r_{\theta_h} = 0$, r is locally strictly concave in θ_h, and anywhere that $r_{\theta_l} = r_{\theta_h} = 0$, r is locally strictly concave in (θ_l, θ_h). Some intuition comes from (29), where we show that after some cancellations, $r_{\theta_h\theta_h}$ has the same sign as

$$\pi_{aa}(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v}) (\gamma(\theta_h, \tilde{\kappa}))_{\theta_h} \left(a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa})\right) + \pi_a(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v}) a_\theta^{-n}(\theta_h),$$

both terms of which are negative at an optimum. The first term reflects that as θ_h increases, effort is distorted further above the efficient level, while the second term reflects that as the action of the opponent gets steeper, the rate at which the firm must distort effort to move θ_h increases.

The proof from here follows the topographical intuition from the introduction. For each θ_h, let $\Theta(\theta_h)$ be the interval of θ_l such that $(\theta_l, \theta_h) \in \Theta \cup A$, so that for $\theta_h' > \theta_T$, $\Theta(\theta_h') = \{\underline{\theta}\}$. Define $\psi(\theta_h) = \max_{\theta_l \in \Theta(\theta_h)} r(\theta_l, \theta_h)$, so that we begin by maximizing r moving east-west. Let D be the set of θ_h such that $\psi > 0$. Fix $\theta_h \in D$ with $\theta_h < \theta_T$. In Lemma 16 we show that $r(\cdot, \theta_h)$ is strictly single-peaked where it is positive and has a unique maximum $\lambda(\theta_h)$. One implication of this is that any local minima are under water. The proof rests on Lemma 12, but accounts for the fact that our terrain is kinked at the boundaries where an opponent changes.

The locus $(\lambda(\cdot), \cdot)$ is the path described in the introduction. We show (Lemma 17) that λ is continuous, and hence so is ψ. We also show (Lemma 18) that D is an interval. We show that the path never runs along L_N, because at any point on L_N, profits are strictly decreasing in θ_l. The path may run along L_S, but we show that where λ is on L_S, ψ is strictly increasing, where the intuition is that on L_S, the firm is better off to strictly increase θ_h, and is also benefited by the fact that L_S is less binding as θ_h increases.

So, consider any $\hat{\theta}_h$ such that $\lambda(\hat{\theta}_h)$ is in the interior of $\Theta(\hat{\theta}_h)$. We show (Lemma 20) that the left and right derivatives of ψ at $\hat{\theta}_h$ and the left and right partial derivatives of r with respect to θ_h at $(\lambda(\hat{\theta}_h), \hat{\theta}_h)$ agree. Given that $\lambda(\theta_h)$ maximizes $r(\cdot, \theta_h)$, this follows from the Envelope Theorem. The proof again deals with kinks in v^{-n} at either θ_h or $\lambda(\theta_h)$.

Using Lemma 18 and Lemma 20, we show (Lemma 21) that ψ is strictly single-peaked–and thus has a unique maximum–on the interval D, which is to say, as one walks northward along the path. This uses the concavity properties already established for r, with the usual complexities at kink points. Finally, we show (Lemma 22) that if θ_h^* is the unique maximizer of ψ, then $(\lambda(\theta_h^*), \theta_h^*)$ is the unique maximizer of r.

Assume that instead of hitting R's western boundary, L_S instead hits R's northern boundary at $(\tilde{\theta}_T, \bar{\theta})$. Then, we can argue as before that any optimum of r occurs either in Θ, with both constraints binding, or on the segment of the northern boundary of R with $\theta_l \leq \tilde{\theta}_T$ with the constraint at $\bar{\theta}$ slack. We can thus perform the same analysis as above, but exchange the roles of θ_l and θ_h, so that one defines $\tilde{\lambda}(\theta_l)$ by first maximizing along north-south slices where θ_l is held constant, and then walks eastward along the path defined by $\tilde{\lambda}$.

6.6 Sufficiency

Let us now discuss sufficiency. Fix $\hat{s}$ satisfying PS, IO, and OB. We wish to show that there is a strategy profile s that is essentially equivalent to $\hat{s}$ and is a Nash equilibrium. The key is that PS, IO and OB imply that, for each n, $\hat{s}^n$ corresponds to a critical point of r. By the previous section (and in particular, by Lemma 22), (θ_l, θ_h) uniquely maximizes r. But then, by Corollary

2, s^n is a best-response to s^{-n}.

The first step is to modify each $\hat{s}^n$ outside of $[\theta_l, \theta_h]$ so as to satisfy $C1$ and $C2$ there as well, so that the results of the previous section apply. We do this while maintaining continuity of actions, and hence, OB is unaffected. Let s be the strategy profile constructed in this way. Consider first $n \notin \{1, N\}$. Since n earns positive profits, and since by IO, the associated κ is in $(H(\theta_l), H(\theta_h))$, Lemma 16 applies and $r(\cdot, \theta_h)$ is single-peaked where it is positive. But then, since $r_{\theta_l}(\theta_l, \theta_h) = 0$ by OB, we must have $\theta_l = \lambda(\theta_h)$, and so, $r_{\theta_h}(\lambda(\theta_h), \theta_h) = \psi_{\theta_h}(\theta_h) = 0$ by Lemma 20. Thus by Lemma 21 $\theta_h = \theta_h^*$, and so s^n is a best response to s^{-n}. The argument for $n = 1$ and N is similar.

6.7 Existence

Let us turn to existence. Recall that in general, Π can be discontinuous, and that there is no reason to believe that the set of best responses is convex. Our plan is to restrict the strategy space so that continuity and convexity of best-responses hold, and to show that the equilibrium of the restricted game is an equilibrium of the original game.

To begin, we need a convex and compact set of strategies. Let η be a bound on both the slope and value of any γ strategy with κ in $[0, 1]$. Impose ($C3$) that action profiles have slope bounded by η. Choose β small enough that if surplus at $\bar{\theta}$ is strictly less than β, then (9) is guaranteed to fail, and impose ($C4$) that the surplus function gives surplus at least β at $\bar{\theta}$. Let S_R^n be the subset of S^n such that $C1$–$C4$ hold, with S_R and S_R^{-n} defined in the usual way.

We first show (Lemma 23) that if other firms choose from S_R^{-n}, then Firm n has a best response in S_R^n. The idea is that every best response is a γ strategy where it is single-dominant, and that η and β were chosen to not bind for such strategies, so that $C3$ and $C4$ are non-binding. Satisfying $C1$ and $C2$ involves inessential modification of the strategy outside of $[\theta_l, \theta_h]$.

Given Lemma 23, it is enough to show that $(S_R^n, \Pi^n)_{n=1}^N$ has an equilibrium. We first establish that S_R^n, and hence S_R, is a Banach space with a norm yielding continuous payoffs. The key to continuity is that $C1$, $C2$, and stacking imply that v^n and v^{n+1} strictly single cross, and hence boundaries move continuously as strategies vary. Compactness and convexity follow since the relevant action profiles are equicontinuous by $C3$, and since $C1$–$C4$ can be phrased as a collection of weak inequalities. Since payoffs are continuous on S_R, BR_R^n has a closed and non-empty valued graph. Finally, to show that $BR_R^n(s^{-n})$ is convex, observe that for any s^{-n}, Section 6.5 implies that any two best responses are essentially equivalent–they hire the same set of workers and give the same surplus to those workers. But then, their convex combination is essentially equivalent to either of them, and so is also a best response. We thus have all the conditions to apply the Kakutani-Fan-Glicksberg Theorem, and hence a Nash equilibrium exists.

7 Conclusion

We analyze an oligopoly market with heterogeneous vertically-differentiated firms and workers with privately known ability. The model is a natural extension to an oligopolistic setting of the ubiquitous principal-agent problem in Mussa and Rosen (1978) and Maskin and Riley (1984). Firms post menus to both screen workers and attract the right pool of applicants. Our analysis uncovers several insights regarding sorting, distortions, and gaps in productivity across firms. We examine the model's competitive limit. Contrary to the monopoly model, asymmetric information can help firms and hurt workers. Finally, we show that under enough firm heterogeneity a simple set of conditions is sufficient for a strategy profile to be an equilibrium, and an equilibrium exists.

There are many extensions of our analysis that are worth pursuing, some for completeness and some more drastic. First is to allow for more general disutility of effort. We conjecture that this will primarily present technical complications. Second is to extend the existence and sufficiency results to the case where firms are less vertically differentiated, so that stacking does not hold. Our existing proof relied hard on stacking to establish continuity. Third is to extend the model to allow both horizontal and vertical differentiation. Fourth, a pressing but challenging extension is to allow for common values and risk-averse workers, so as to apply the framework to insurance markets. Finally, it would be of great interest to incorporate moral hazard in a nontrivial way.

8 Appendix A: Proofs for Sections 4–5

8.1 Proof of PP

We begin with a preliminary result. It shows that there is zero probability that a firm hires a worker on whom it strictly loses money, and that among each firm's best responses is always a menu in which every offer, whether accepted with positive probability or not, is profitable.

Proposition 3 *Fix n, s^{-n}, and $s^n = (\alpha, v)$. Let $P \equiv \{\theta | \pi(\theta, \alpha, v) \geq 0\}$. Then, there is $(\hat{\alpha}, \hat{v})$ with $\pi(\cdot, \hat{\alpha}, \hat{v}) \geq 0$ that agrees on P with (α, v). If (α, v) is a best response to s^{-n}, then $\pi(\theta, \alpha, v) \geq 0$ for almost all θ where $\varphi > 0$.*

Proof The idea is simply to remove all menu items for which θ is not in P. Let us first show that P can be taken to be closed. Formally, fix n, and let $G(\theta, v)$ be the subdifferential to v at θ. Since v is convex, G is singleton-valued almost everywhere, and every selection from G is increasing. Thus, since G is compact-valued, it is wlog to assume that $\alpha(\theta) \in \arg\max_{a \in G(\theta,v)} \pi(\theta, a, v)$ for all θ. But then, since G is upper hemicontinuous in θ, $\pi(\cdot, \alpha, v)$ is upper semicontinuous (Aliprantis and Border (2006), Lemma 17.30, p. 569), and so $P \equiv \{\theta | \pi(\theta, \alpha, v) \geq 0\}$ is a closed subset of $[\underline{\theta}, \overline{\theta}]$, and hence compact.

Now let's build the menu that results when menu items with θ not in P are removed. For each $\theta' \in [\underline{\theta}, \bar{\theta}]$, let $v_L(\cdot, \theta')$ be the line given by $v_L(\theta, \theta') = v(\theta') + (\theta - \theta')\alpha(\theta')$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. Note that $v_L(\theta, \theta) = v(\theta)$, that since v is convex, $v_L(\cdot, \theta')$ lies below v for each θ', and that along $v_L(\cdot, \theta')$, the profits to the firm are constant. If P is empty, set $(\alpha, v) = (\alpha_*, v_*)$, and we are done. If P is non-empty, define $\hat{v}(\theta) = \max_{\theta' \in P} v_L(\theta, \theta')$. Then, $\hat{v}$, which is the maximum of a set of lines, is convex, with $\hat{v} = v$ on P (using that $v_L(\theta, \theta) = v(\theta)$) and $\hat{v} \leq v$ (since each $v_L(\cdot, \theta')$ lies below v). Let $\hat{\alpha}$ be a selection from $G(\cdot, \hat{v})$, where we can take $\hat{\alpha} = \alpha$ on P, and where at any $\theta \notin P$, we can take $\hat{\alpha}(\theta) = \alpha(\theta')$ for some $\theta' \in \arg\max_{\theta' \in P} v_L(\theta, \theta')$. Then by using $(\hat{\alpha}, \hat{v})$, the firm implements the same action on P at the same profit as before (the types in P have no new deviations available), and the firm earns positive profits on any other worker, since that worker either leaves or, if hired, is now imitating a worker in P.

Note finally that if (α, v) is a best response to s^{-n}, and $\pi(\theta, \alpha, v) < 0$ for some positive measure set of θ where $\varphi > 0$, then $(\hat{\alpha}, \hat{v})$ gives strictly higher profits than (α, v), a contradiction. □

Corollary 3 *Each firm earns strictly positive profits in equilibrium.*

Proof By assumption there is θ_*^n such that $v_*^n(\theta_*^n) > v_*^{-n}(\theta_*^n)$, and so, by continuity, $v_*^n(\theta) > v_*^{-n}(\theta)$ for all θ in some interval I around θ_*^n. Assume that on a positive measure set of I, $v^{-n}(\theta) \geq v_*^n(\theta)$. Then, since $v_*^n(\theta) > v_*^{-n}(\theta)$ on I, either some firm other than n is winning with positive probability and is losing money, or n is winning having offered surplus $v^n(\theta) > v^{-n}(\theta) \geq v_*^n(\theta)$. (Note that if n offers $v^{-n}(\theta)$, then firms other than n win with positive probability since ties are broken equiprobably.) Either case violates Proposition 3. But then, for ε sufficiently small but positive, the strategy of offering all types surplus $v_*^n(\theta) - \varepsilon$ and action $\alpha_*^n(\theta)$ earns at least ε on a positive measure set of types. Hence, n must earn strictly positive profits in equilibrium. □

This proof used in an essential way that for each n, φ^n is strictly positive where $v^n = v^{-n}$. To see this, assume that workers, if indifferent, sort themselves to a firm that makes the most money on them. Then, a zero-profit equilibrium of the game is that each firm offers the same menu (α_*, v_*), where (recall) v_* is the most surplus any firm can offer without losing money, and where α_* is the associated efficient action for a relevant firm. Since workers sort efficiently, no firm loses money, while given that other firms are offering v_*, no strictly profitable deviation exists.

8.2 Proof of NP

We now formalize and establish property NP. We show that in any optimal strategy, the firm hires with probability 1 on the set of types, $Z_{\geq}$ which it sometimes hires and on whom it makes strictly positive profits, and that the set of types $Z_{<}$ that it is does not hire but where it could strictly profitably imitate its competitors is empty.

Proposition 4 *Let (α^n, v^n) be optimal given s^{-n}, and define $Z_{\geq} = \{\theta | v^n \geq v^{-n}$ and $\pi^n(\theta, \alpha^n, v^n) > 0\}$ as the set of types where n wins at least sometimes, and makes strict profits, and $Z_{<} = \{\theta | v^n < v^{-n}$ and $\pi^n(\theta, a^{-n}, v^{-n}) > 0\}$ as the set of types where n never wins, but could profitably imitate the incumbent. Then, $\int_{Z_{\geq}} (1 - \varphi(\theta, s))\, h(\theta) d\theta = 0$, and $Z_{<}$ is empty*

Proof Essentially, the firm can first imitate (a^{-n}, v^{-n}) anywhere that $v^{-n} > v^n$, then purge any unprofitable menus items using Proposition 3, and then add ε, ensuring a hire wherever it is profitable. Formally, let $\tilde{v} = \max\{v^n, v^{-n}\}$ and let $\tilde{\alpha} = \alpha^n$ where $v^n \geq v^{-n}$ and $\tilde{\alpha} = a^{-n}$ where $v^n < v^{-n}$. Using Proposition 3, define a new menu $(\hat{\alpha}, \hat{v})$ that agrees with $(\tilde{\alpha}, \tilde{v})$ anywhere that $\pi^n(\theta, \tilde{\alpha}, \tilde{v}) \geq 0$ and satisfies $\pi^n(\theta, \hat{\alpha}, \hat{v}) \geq 0$ for all θ. Note that $\hat{v} \geq v^{-n}$ wherever $\pi^n(\theta, \alpha^n, v^n) \geq 0$. Thus (regardless of the tie-breaking rule), the menu $(\hat{\alpha}, \hat{v} + \varepsilon)$, $\varepsilon > 0$, earns at least

$$\int_{Z_{\geq}} \pi^n(\theta, \hat{\alpha}, \hat{v}) h(\theta) d\theta + \int_{Z_{<}} \pi^n(\theta, \hat{\alpha}, \hat{v}) h(\theta) d\theta - \varepsilon = \int_{Z_{\geq}} \pi^n(\theta, \alpha^n, v^n) h(\theta) d\theta + \int_{Z_{<}} \pi^n(\theta, a^{-n}, v^{-n}) h(\theta) d\theta - \varepsilon.$$

Hence, since $\varepsilon > 0$ is arbitrary and (α^n, v^n) is optimal, we must have

$$\int_{Z_{\geq}} \pi^n(\theta, \alpha^n, v^n) \varphi^n(\theta, s) h(\theta) d\theta \geq \int_{Z_{\geq}} \pi^n(\theta, \alpha^n, v^n) h(\theta) d\theta + \int_{Z_{<}} \pi^n(\theta, a^{-n}, v^{-n}) h(\theta) d\theta,$$

and thus

$$\int_{Z_{\geq}} \pi^n(\theta, \alpha^n, v^n)(1 - \varphi^n(\theta, s)) h(\theta) d\theta + \int_{Z_{<}} \pi^n(\theta, a^{-n}, v^{-n}) h(\theta) d\theta \leq 0,$$

which, given the definitions of $Z_{\geq}$ and $Z_{<}$, only occurs if $Z_{<}$ is empty and $\int_{Z_{\geq}} (1 - \varphi^n)\, h = 0$. $\square$

8.3 Proof of PS

Let us first prove that any Nash equilibrium (with or without NEO) satisfies a condition slightly weaker than PS. Say that s has *quasi-positive sorting (QPS)* if it satisfies the conditions for PS except that each condition on φ is allowed to fail on a zero-measure subset.

Proposition 5 *Every Nash equilibrium has QPS.*

Proof Let $n' > n$, let $\theta_{\text{inf}}^{n'}$ be the infimum of the support of $\varphi^{n'}$ and let θ_{sup}^{n} be the supremum of the support of φ^n. We will show that the only way that $\theta_{\text{inf}}^{n'} < \theta_{\text{sup}}^{n}$ can hold is if $n = n + 1$, and the two firms are tied at zero profits on $(\theta_{\text{inf}}^{n'}, \theta_{\text{sup}}^{n})$. The core of the proof is to exploit that B^n is strictly super-modular in n and a.

Assume that $\theta_{\text{inf}}^{n'} < \theta_{\text{sup}}^{n}$. Conditional on $\varphi^{n'}(\theta, s) > 0$, with probability one $\pi^{n'}(\theta, \alpha^{n'}, v^{n'}) \geq 0$ by Proposition 3 and $\pi^n(\theta, \alpha^{n'}, v^{n'}) \leq 0$ by Proposition 4. Hence, for any $\varepsilon \in (0, (\theta_{\text{sup}}^{n} - \theta_{\text{inf}}^{n'})/2)$

there is $\theta_1 \in [\theta^{n'}_{\inf}, \theta^{n'}_{\inf} + \varepsilon]$ where $\varphi^{n'}(\theta_1) > 0$ and

$$\pi^{n'}(\theta_1, \alpha^{n'}, v^{n'}) \geq 0 \geq \pi^n(\theta_1, \alpha^{n'}, v^{n'}), \tag{14}$$

and similarly, there is $\theta_2 \in [\theta^n_{\sup} - \varepsilon, \theta^n_{\sup}]$ where $\varphi^n(\theta_2) > 0$ and

$$\pi^n(\theta_2, \alpha^n, v^n) \geq 0 \geq \pi^{n'}(\theta_2, \alpha^n, v^n). \tag{15}$$

By incentive compatibility, since $\theta_2 > \theta_1$ and since $\varphi^{n'}(\theta_1) > 0$ and $\varphi^n(\theta_2) > 0$, it must be that $\alpha^n(\theta_2) \geq \alpha^{n'}(\theta_1)$. Adding (14) and (15) and cancelling common terms,

$$B^{n'}(\alpha^{n'}(\theta_1)) + B^n(\alpha^n(\theta_2)) \geq B^n(\alpha^{n'}(\theta_1)) + B^{n'}(\alpha^n(\theta_2)).$$

Since $B^n(a)$ is strictly supermodular, $\alpha^{n'}(\theta_1) = \alpha^n(\theta_2) \equiv \tilde{a}$, and so, by incentive compatibility, and since ε was arbitrary, $\alpha^{n'}(\theta) = \alpha^n(\theta) = \tilde{a}$ for all $\theta \in (\theta^{n'}_{\inf}, \theta^n_{\sup})$. From (14), $B^{n'}(\tilde{a}) \geq B^n(\tilde{a})$, while from (15), $B^{n'}(\tilde{a}) \leq B^n(\tilde{a})$, and so $B^{n'}(\tilde{a}) = B^n(\tilde{a}) \equiv b$. But then, from (14), $\pi^{n'}(\theta_1, \alpha^{n'}, v^{n'}) = 0$, and from (15), $\pi^n(\theta_2, \alpha^n, v^n) = 0$. Finally, on $(\theta^{n'}_{\inf}, \theta^n_{\sup})$, $(\pi(\theta, \alpha, v))_\theta = \pi_a(\theta, \alpha, v)\alpha_\theta(\theta) = 0$, using $-c_\theta(\alpha(\theta), \theta) = \alpha(\theta) = v_\theta(\theta)$. Hence $\pi^n = \pi^{n'} = 0$ on $(\theta^{n'}_{\inf}, \theta^n_{\sup})$.

Now let us show that $n' = n + 1$. Assume that $n' \neq n + 1$, and let $n < n'' < n'$. Assume first that $B^{n''}(\tilde{a}) \leq b = B^n(\tilde{a})$. Then since $n'' > n$ and $B^n(a)$ is strictly supermodular, $B^{n''}(a) < B^n(a)$ for all $a < \tilde{a}$, and similarly, $B^{n''}(a) < B^{n'}(a)$ for all $a > \tilde{a}$, contradicting that $B^{n''}$ is somewhere uniquely maximal. Thus $B^{n''}(\tilde{a}) > b$, and so $\pi^{n''}(\theta, \tilde{a}, v^{-n}) > 0$ on $(\theta^{n'}_{\inf}, \theta^n_{\sup})$, which contradicts Proposition 4 since by definition of $\theta^{n'}_{\inf}$ and $\theta^n_{\sup}$, $\int_{\theta^{n'}_{\inf}}^{\theta^n_{\sup}} (1 - \varphi^{n''})h > 0$. Thus, $n' = n + 1$, and $\tilde{a} = \hat{a}^n$. Letting $\theta^n_h = \theta^{n'}_{\inf}$ and $\theta^{n+1}_l = \theta^n_{\sup}$, we have the claimed structure at ties.

Finally, it must be that $\theta^n_l < \theta^n_h$, since by Corollary 3, n earns strictly positive expected profit, but on each type above θ^n_h or below θ^n_l either loses for sure or ties but earns 0. □

Corollary 4 *Every Nash Equilibrium that satisfies NEO has PS.*

Proof Assume that for some $n' > n$, and for some $\hat{\theta} \in (\theta_l, \theta_h)$, $v^{n'} = v^n$. Then, since by *NEO*, $\alpha^{n'} \geq \alpha^n$, and hence $v^{n'}(\theta) - v^n(\theta)$ is increasing, $v^{n'} \geq v^n$ everywhere on $[\hat{\theta}, \theta_h]$, contradicting that n wins with probability one conditional on $\theta \in (\theta_l, \theta_h)$. □

Next, let us show that any best response must be continuous anywhere that it is "active."

Lemma 5 *Fix n, s^{-n}, and $\hat{s} = (\hat{\alpha}, \hat{v})$. If $\hat{s}$ is a best-response, then $\hat{\alpha}$ must be continuous on any open interval where $v^n \geq v^{-n}$.*

Proof Essentially, because π is strictly concave in a, any jump in $\hat{\alpha}$ creates an opportunity for a strictly profitable perturbation. In particular, let θ_J be a point on some open interval where $v^n \geq v^{-n}$ where $\hat{\alpha}$ jumps from $\underline{a}$ to $\bar{a}$. Raise $\hat{\alpha}$ by q on $[\theta_J - \varepsilon, \theta_J)$ and lower it by q on $[\theta_J, \theta_J + \varepsilon]$ where for ε and q small enough, monotonicity is respected. This raises surplus slightly on $(\theta_J - \varepsilon, \theta_J + \varepsilon)$ (by an amount at most $q\varepsilon$), but otherwise does not affect v. The perturbed strategy hires with probability one on $(\theta_J - \varepsilon, \theta_J + \varepsilon)$, and any new worker hired by this perturbation is profitable, since by Proposition 3, $\hat{s}$ loses money nowhere. We claim that because π is strictly concave in a, this perturbation is strictly profitable for sufficiently small ε and q, contradicting the optimality of $\hat{s}$.

To see this formally, let $\hat{s}(q, \varepsilon) = (\alpha(\cdot, q, \varepsilon), v(\cdot, q, \varepsilon))$ be the resultant menu, and note that for $\theta \in [\theta_J - \varepsilon, \theta_J)$, $\alpha_q(\theta, q, \varepsilon) = 1$, and $v_q(\theta, q, \varepsilon) \leq \varepsilon$. Hence,

$$\frac{\partial}{\partial q}\pi(\theta, s(q, \varepsilon)) \geq \pi_a(\theta, s(q, \varepsilon)) - \varepsilon \geq \pi_a(\theta, \underline{a} + q, v(q)) - \varepsilon,$$

since π is concave in a. Similarly, for $[\theta_J, \theta_J + \varepsilon]$

$$\frac{\partial}{\partial q}\pi(\theta, s(q, \varepsilon)) \geq -\pi_a(\theta, s(q, \varepsilon)) - \varepsilon \geq -\pi_a(\theta, \bar{a} - q, v(q)) - \varepsilon.$$

Hence, recalling that $=_s$ means "has strictly the same sign as,"

$$\begin{aligned}
&\frac{\partial}{\partial q}\Pi\left(\hat{s}(q, \varepsilon), s^{-n}\right) \\
&\geq \int_{\theta_J - \varepsilon}^{\theta_J} (\pi_a(\theta, \underline{a} + q, v(q)) - \varepsilon) h(\theta) d\theta + \int_{\theta_J}^{\theta_J + \varepsilon} (-\pi_a(\theta, \underline{a} + q, v(q)) - \varepsilon) h(\theta) d\theta \\
&= \varepsilon\left[\left(\pi_a\left(\theta', \underline{a} + q, v(q)\right) - \varepsilon\right) h\left(\theta'\right) - \left(\pi_a\left(\theta'', \underline{a} + q, v(q)\right) - \varepsilon\right) h\left(\theta''\right)\right] \\
&\underset{s}{=} \left(\pi_a\left(\theta', \underline{a} + q, v(q)\right) - \varepsilon\right) h\left(\theta'\right) - \left(\pi_a\left(\theta'', \underline{a} + q, v(q)\right) - \varepsilon\right) h\left(\theta''\right) \\
&\cong (\pi_a(\theta_J, \underline{a} + q, v(q)) - \pi_a(\theta_J, \underline{a} + q, v(q))) h(\theta_J) \\
&> 0
\end{aligned}$$

where the first equality uses the Mean Value Theorem for some $\theta' \in [\theta_J - \varepsilon, \theta_J]$ and $\theta'' \in [\theta_J, \theta_J + \varepsilon]$, where the approximation is arbitrarily good when ε is small, and where the last inequality holds for q small. But then, for ε and q small, $\frac{\partial}{\partial q}\Pi(\hat{s}(q, \varepsilon), s^{-n}) > 0$, and we are done. □

8.4 Proof of *IO*

We begin with two preliminary lemmas. The first one is central in showing that strategies of the γ form are monotone.

Lemma 6 *Let $\kappa \in [0,1]$. Then,*

$$\left(\frac{\kappa - H(\theta)}{h(\theta)}\right)_\theta = -1 - \frac{(\kappa - H(\theta))\, h'(\theta)}{h^2\,(\theta)} \leq 0. \tag{16}$$

Proof Assume first that $h'\,(\theta) \leq 0$. Then,

$$\frac{(\kappa - H(\theta))\, h'(\theta)}{h^2(\theta)} \geq \frac{(1 - H(\theta))h'(\theta)}{h^2(\theta)} \geq -1,$$

since $1 - H$ is log-concave. If $h'(\theta) > 0$, then the result follows since H is log-concave. □

Our next lemma re-expresses profits of the firm in a useful and standard way.

Lemma 7 *Fix n, and for any feasible α and v, define*

$$M(\theta, \alpha, v) = B(\alpha(\theta)) - c(\alpha(\theta), \theta) - v(\theta_l) - \alpha(\theta)\frac{H(\theta_h) - H(\theta)}{h(\theta)}. \tag{17}$$

Then,

$$\int_{\theta_l}^{\theta_h} \pi(\theta, \alpha, v)h(\theta)d\theta = \int_{\theta_l}^{\theta_h} M(\theta, \alpha, v)h(\theta)d\theta. \tag{18}$$

Proof Note first that for any α and v,

$$\int_{\theta_l}^{\theta_h} \pi(\theta, \alpha, v)h(\theta)d\theta = \int_{\theta_l}^{\theta_h} \left(B(\alpha(\theta)) - c(\alpha(\theta), \theta) - v(\theta_l) - \int_{\theta_l}^{\theta} \alpha(\tau)d\tau\right) h(\theta)d\theta,$$

and that, integrating by parts,

$$\int_{\theta_l}^{\theta_h} \left(\int_{\theta_l}^{\theta} \alpha\,(\tau)\, d\tau\right) h(\theta)d\theta = \int_{\theta_l}^{\theta_h} \alpha(\theta)(H(\theta_h) - H(\theta))d\theta.$$

Substituting and rearranging yields (18). □

We now prove Proposition 1, which derives the solution to the relaxed problem $\mathcal{P}(\theta_l, \theta_h)$.

Proof of Proposition 1 Existence is standard and uniqueness follows since the set of feasible strategies is convex, and the objective function is strictly concave (since $\pi\,(\theta, a, v)$ is strictly concave in a and linear in v for each θ). Fix (θ_l, θ_h), fix the optimum $\tilde{s} = (\tilde{\alpha}, \tilde{v})$, and define

$$\xi(\theta) = \pi_a(\theta, \tilde{s})h(\theta) + H(\theta).$$

Step 1 Let us show that ξ is constant on (θ_l, θ_h) at some value κ_o and thus, rearranging, that

$$\pi_a\,(\theta, \tilde{s}) = \frac{\kappa_o - H\,(\theta)}{h\,(\theta)}.$$

To see the idea, choose any two points $\theta'' > \theta'$ in (θ_l, θ_h). We will consider perturbations which subtract (or add) a small amount from the action schedule near θ', and replace it near θ''. We can do this without worrying about monotonicity, since this is the relaxed problem. This perturbation has cost $\pi_a h$ near θ', benefit $\pi_a h$ near θ'', and benefit $H(\theta'') - H(\theta')$ because v is lowered between θ' and θ''. But then, setting net benefit equal to zero, we have

$$-\pi_a(\theta', \tilde{s})\, h(\theta') + H(\theta'') - H(\theta') + \pi_a(\theta'', \tilde{s})\, h(\theta'') = 0,$$

or, rearranging, $\xi(\theta') = \xi(\theta'')$. Setting $\kappa_o = \xi(\theta)$ for any $\theta \in (\theta_l, \theta_h)$, we are done.

To formalize this, fix $0 < \varepsilon < \frac{1}{2}\min\{\theta'' - \theta', \theta' - \theta_l, \theta_h - \theta''\}$. Define $\hat{\alpha}(\cdot, y, \varepsilon)$ to be $\tilde{\alpha} - y/2\varepsilon$ on $[\theta' - \varepsilon, \theta' + \varepsilon]$, $\tilde{\alpha} + y/2\varepsilon$ on $[\theta'' - \varepsilon, \theta'' + \varepsilon]$, and $\tilde{\alpha}$ elsewhere, and define

$$\hat{v}(\theta, y, \varepsilon) = \tilde{v}(\theta_l) + \int_{\theta_l}^{\theta} \hat{\alpha}(\tau, y, \varepsilon) d\tau,$$

noting that $\hat{v}(\theta_h, y, \varepsilon) = \tilde{v}(\theta_h)$, and so for each y, $\hat{s}(y, \varepsilon) = (\hat{\alpha}(\cdot, y, \varepsilon), \hat{v}(\cdot, y, \varepsilon))$ is feasible in $\mathcal{P}(\theta_l, \theta_h)$. Note that $\hat{v}_y(\theta, y, \varepsilon) = -1$ on $[\theta' + \varepsilon, \theta'' - \varepsilon]$, and $\hat{v}_y(\theta, y, \varepsilon) \in [-1, 0]$ on $[\theta' - \varepsilon, \theta' + \varepsilon]$ and $[\theta'' - \varepsilon, \theta'' + \varepsilon]$.

Let profits of this perturbation as a function of y and ε be $j(y, \varepsilon) = \int_{\theta_l}^{\theta_h} \pi(\theta, \hat{s}(y, \varepsilon)) h(\theta) d\theta$. Then, since $\pi_v = -1$,

$$\begin{aligned} j_y(y, \varepsilon) &= \int_{\theta' - \varepsilon,}^{\theta' + \varepsilon} (-\pi_a(\theta, \hat{s}(y, \varepsilon))\frac{1}{2\varepsilon} - \hat{v}_y(\theta, y, \varepsilon)) h(\theta) d\theta + \int_{\theta' + \varepsilon}^{\theta'' - \varepsilon} h(\theta) d\theta \\ &\quad + \int_{\theta'' - \varepsilon,}^{\theta'' + \varepsilon} (\pi_a(\theta, \hat{s}(y, \varepsilon))\frac{1}{2\varepsilon} - \hat{v}_y(\theta, y, \varepsilon)) h(\theta) d\theta, \end{aligned}$$

where between $\theta' + \varepsilon$ and $\theta'' - \varepsilon$ we use $\hat{\alpha}_y = 0$ and $\hat{v}_y = -1$. Note that $\hat{s}(0, \varepsilon) = (\tilde{\alpha}, \tilde{v})$. Hence, evaluating $j_y(y, \varepsilon)$ at $y = 0$, and using the Mean Value Theorem, there is $\tau' \in [\theta' - \varepsilon, \theta' + \varepsilon]$ and $\tau'' \in [\theta'' - \varepsilon, \theta'' + \varepsilon]$ such that

$$\begin{aligned} j_y(0, \varepsilon) &= 2\varepsilon \left(-\pi_a(\tau', \tilde{\alpha}, \tilde{v})\frac{1}{2\varepsilon} - \hat{v}_y(\tau', 0, \varepsilon) \right) h(\tau') + \big(H(\theta'' - \varepsilon) - H(\theta' + \varepsilon)\big) \\ &\quad + 2\varepsilon \left(\pi_a(\tau'', \tilde{\alpha}, \tilde{v})\frac{1}{2\varepsilon} - \hat{v}_y(\tau'', 0, \varepsilon) \right) h(\tau''). \end{aligned}$$

But then, since $\hat{v}_y(\tau', 0, \varepsilon)$ and $\hat{v}_y(\tau'', 0, \varepsilon)$ are bounded,

$$\lim_{\varepsilon \to 0} j_y(0, \varepsilon) = -\pi_a(\theta', \tilde{\alpha}, \tilde{v}) h(\theta') + H(\theta'') - H(\theta') + \pi_a(\theta'', \tilde{\alpha}, \tilde{v}) h(\theta'') = \xi(\theta'') - \xi(\theta'),$$

and so, if $\xi(\theta'') - \xi(\theta') \neq 0$, then for ε sufficiently small, $j_y(0, \varepsilon) \neq 0$, and the firm has a profitable deviation, a contradiction.

Step 2 Let us next show that if one fixes surplus to equal $\tilde{v}(\theta_l)$ at θ_l, and then varies κ, ignoring (6), then profits are single-peaked at $\kappa = H(\theta_h)$. Similarly, if one fixes surplus to equal $\tilde{v}(\theta_h)$ at θ_h, and then varies κ, ignoring (5), then profits are single-peaked at $\kappa = H(\theta_l)$.

To formalize this, let $v_l(\theta,\kappa) = \tilde{v}(\theta_l) + \int_{\theta_l}^{\theta} \gamma(\tau,\kappa)d\tau$, and let $s_l(\kappa) = (\gamma(\cdot,\kappa), v_l(\cdot,\kappa))$. Since $v_l(\theta_l,\kappa) = \tilde{v}(\theta_l)$ and so is independent of κ, it follows from (17) that on (θ_l,θ_h),

$$\begin{aligned} \frac{d}{d\kappa}M(\theta, s_l(\kappa)) &= \left(\pi_a(\theta, s_l(\kappa)) - \frac{H(\theta_h)-H(\theta)}{h(\theta)}\right)\gamma_\kappa(\theta,\kappa) \\ &= \left(\frac{\kappa - H(\theta_h)}{h(\theta)}\right)\gamma_\kappa(\theta,\kappa) \underset{s}{=} -(\kappa - H(\theta_h)), \end{aligned} \tag{19}$$

since $\gamma_\kappa < 0$. But then, letting $Y_l(\kappa) \equiv \int_{\theta_l}^{\theta_h} \pi(\theta, s_l(\kappa))h(\theta)d\theta$, by Lemma 7, $dY_l(\kappa)/d\kappa \underset{s}{=} -(\kappa - H(\theta_h))$, and so $Y_l(\kappa)$ is strictly single-peaked at $\kappa = H(\theta_h)$.

Similarly, if we define $v_h(\theta,\kappa) = \tilde{v}(\theta_h) - \int_{\theta}^{\theta_h} \gamma(\tau,\kappa)d\tau$, then $Y_h(\kappa) \equiv \int_{\theta_l}^{\theta_h} \pi(\theta, \gamma(\cdot,\kappa), v_h(\cdot,\kappa))h(\theta)d\theta$ is strictly single-peaked in κ with maximum at $\kappa = H(\theta_l)$ where to show this, one integrates

$$\int_{\theta_l}^{\theta_h} \pi(\theta,\alpha,v)h(\theta)d\theta = \int_{\theta_l}^{\theta_h}\left(B(\alpha(\theta)) - c(\alpha(\theta),\theta) - v(\theta_h) + \int_{\theta}^{\theta_h}\alpha(\tau)d\tau\right)h(\theta)d\theta$$

by parts to arrive at an analogue to M.

Step 3 Finally, let us use Step 2 to show that $\kappa_o = \tilde{\kappa}(\theta_l,\theta_h)$. Note that one of (5) and (6) must bind, otherwise reducing $\tilde{v}$ by a small positive constant (holding fixed $\tilde{\alpha}$) is profitable. Assume that $\tilde{v}(\theta_h) > v^{-n}(\theta_h)$. Then, $s_l(\kappa)$ is feasible for κ on a neighborhood of κ_o, and so, since Y_l is strictly single-peaked with maximum at $H(\theta_h)$ we must have $\kappa_o = H(\theta_h)$. Since $\tilde{v}(\theta_l) = v^{-n}(\theta_l)$, and since $(\tilde{\alpha},\tilde{v}) = (\gamma(\cdot,\kappa_o),\tilde{v})$ is feasible, we have

$$\tilde{v}(\theta_h) = v^{-n}(\theta_l) + \int_{\theta_l}^{\theta_h} \gamma(\tau, H(\theta_h))d\tau > v^{-n}(\theta_h),$$

and so $z(\theta_l,\theta_h,H(\theta_h)) < 0$, and thus by definition of $\tilde{\kappa}$, we have $\tilde{\kappa}(\theta_l,\theta_h) = H(\theta_h)$ as well, so that $\kappa_o = \tilde{\kappa}(\theta_l,\theta_h)$. Similarly, if $\tilde{v}(\theta_l) > v^{-n}(\theta_l)$ then, using Y_h, we must have $\kappa_o = H(\theta_l) = \tilde{\kappa}(\theta_l,\theta_h)$.

Assume finally that (5) and (6) both bind. Then, by definition, $z(\theta_l,\theta_h,\kappa_o) = 0$. Assume $\kappa_o > H(\theta_h)$. Then,

$$v_l(\theta_h, H(\theta_h)) = \tilde{v}(\theta_l) + \int_{\theta_l}^{\theta_h}\gamma(\tau,H(\theta_h))d\tau > \tilde{v}(\theta_l) + \int_{\theta_l}^{\theta}\gamma(\tau,\kappa_o)d\tau = \tilde{v}(\theta_h) = v^{-n}(\theta_h),$$

so that $s_l(H(\theta_h))$ is feasible, which contradicts the optimality of $(\tilde{\alpha},\tilde{v})$ since Y_l is uniquely maximized at $H(\theta_h)$, and Y_l ignores (6). Hence $\kappa_o \leq H(\theta_h)$. Similarly, $\kappa_o \geq H(\theta_l)$, and thus $\kappa_o \in [H(\theta_l), H(\theta_h)]$, from which $\kappa_o = \tilde{\kappa}(\theta_l,\theta_h)$, again by definition of $\tilde{\kappa}$. □

We now prove that any optimum of the original problem has the form given by Proposition 1.

Proposition 6 *Let s be Nash. Then, for each n, there is $\kappa^n \in [H(\theta_l^n), H(\theta_h^n)]$ such that $\alpha^n = \gamma^n(\cdot, \kappa^n)$ on (θ_l^n, θ_h^n), where $\kappa^1 = 0$, and $\kappa^N = 1$.*

Proof We will show that if (α, v) is not equal to $(\tilde{\alpha}, \tilde{v})$, the optimal solution to the relaxed problem, then we can profitably perturb (α, v) in the direction of $(\tilde{\alpha}, \tilde{v})$.[20] We need this perturbation to respect monotonicity and the fact that workers both within and outside of (θ_l, θ_h) may be affected. This proof would be substantially simpler if all crossings where transversal, but we know this will fail when firms are not very differentiated.

Let $\check{s}(\delta)$ be given by $\check{\alpha}(\cdot, \delta) = (1-\delta)\alpha + \delta\tilde{\alpha}$ and $\check{v}(\cdot, \delta) = (1-\delta)v + \delta\tilde{v}$, so that $\check{s}(0) = (\alpha, v)$ and $\check{s}(1) = (\tilde{\alpha}, \tilde{v})$. The problem with $\check{s}$ is that when crossings are not transversal, $\check{s}(\delta)$ need not hire all of (θ_l, θ_h) even for small δ. So, let $\bar{v} = v^{-n}/2 + v/2$, so that $\bar{v} > v^{-n}$ on (θ_l, θ_h). Now, let $\grave{v}(\cdot, \delta) = \max(\bar{v}, \check{v}(\cdot, \delta))$, let $\grave{\alpha}(\cdot, \delta)$ be a subgradient to $\grave{v}(\cdot, \delta)$, and let $\grave{s}(\delta) = (\grave{\alpha}(\cdot, \delta), \grave{v}(\cdot, \delta))$. By construction, $\grave{s}$ always wins on (θ_l, θ_h), and may hire other workers as well. Also, since on (θ_l, θ_h), $v > v^{-n}$, $\grave{s}(0) = (\alpha, v)$. Finally, let $P(\delta)$ be the set upon which $\grave{s}(\delta)$ is profitable, and construct $\hat{s}(\delta) = (\hat{\alpha}(\cdot, \delta), \hat{v}(\cdot, \delta))$ from $\grave{s}(\delta)$ as in Proposition 3. We then have

$$\begin{aligned}
\Pi(\hat{s}(\delta), s^{-n}) &= \int \pi(\theta, \hat{s}(\delta))\varphi(\theta, \hat{s}(\delta), s^{-n})h(\theta)d\theta \\
&\geq \int_{P(\delta)\cap(\theta_l,\theta_h)} \pi(\theta, \hat{s}(\delta))\varphi(\theta, \hat{s}(\delta), s^{-n})h(\theta)d\theta \\
&= \int_{P(\delta)\cap(\theta_l,\theta_h)} \pi(\theta, \grave{s}(\delta))h(\theta)d\theta \\
&\geq \int_{\theta_l}^{\theta_h} \pi(\theta, \grave{s}(\delta))h(\theta)d\theta.
\end{aligned}$$

The first inequality follows since $\pi(\cdot, \hat{s}(\delta)) \geq 0$, the second equality since $\hat{s}(\delta)$ and $\grave{s}(\delta)$ agree on $P(\delta)$ and $\varphi(\cdot, \grave{s}(\delta)) = 1$ on (θ_l, θ_h), and the second inequality since $\pi(\theta, \grave{s}(\delta)) \leq 0$ outside of $P(\delta)$.

It is thus enough to show that for δ sufficiently small,

$$\int_{\theta_l}^{\theta_h} \pi(\theta, \grave{s}(\delta))h(\theta)d\theta > \int_{\theta_l}^{\theta_h} \pi(\theta, \alpha, v)h(\theta)d\theta,$$

since by PS, $\pi(\theta, \alpha, v)\varphi(\theta, s) = 0$ outside of $[\theta_l, \theta_h]$. Because $\grave{s}(0) = (\alpha, v)$, it is sufficient that

$$\frac{d}{d\delta} \int_{\theta_l}^{\theta_h} \pi(\theta, \grave{s}(\delta))h(\theta)d\theta \bigg|_{\delta=0} > 0.$$

[20]It is not important how $(\tilde{\alpha}, \tilde{v})$ is defined outside of (θ_l, θ_h) so long as monotonicity, continuity of actions, and the integral condition hold.

But,

$$\left.\frac{d}{d\delta}\int_{\theta_l}^{\theta_h}\pi(\theta,\grave{s}(\delta))h(\theta)d\theta\right|_{\delta=0}=\left.\frac{d}{d\delta}\int_{\theta_l}^{\theta_h}\pi(\theta,\check{s}(\delta))h(\theta)d\theta\right|_{\delta=0},$$

since for each $\theta\in(\theta_l,\theta_h)$, $v(\theta)>\bar{v}(\theta)$, and so at $\delta=0$, $(\check{\alpha}(\theta,\delta))_\delta=(\grave{\alpha}(\theta,\delta))_\delta$ and $(\check{v}(\theta,\delta))_\delta=(\grave{v}(\theta,\delta))_\delta$. And, since $(\tilde{\alpha},\tilde{v})$ is the unique solution on (θ_l,θ_h) to the relaxed problem $\mathcal{P}(\theta_l,\theta_h)$, and since $\check{s}(0)=(\alpha,v)$, and so is feasible in $\mathcal{P}(\theta_l,\theta_h)$,

$$\begin{aligned}\int_{\theta_l}^{\theta_h}\pi(\theta,\tilde{\alpha},\tilde{v})h(\theta)d\theta &= \int_{\theta_l}^{\theta_h}\pi(\theta,\check{s}(1))h(\theta)d\theta\\ &> \int_{\theta_l}^{\theta_h}\pi(\theta,\check{s}(0))h(\theta)d\theta=\int_{\theta_l}^{\theta_h}\pi(\theta,\alpha,v)h(\theta)d\theta.\end{aligned}$$

Now, $\check{s}$ is linear in δ, and $\pi(\theta,\cdot,\cdot)$ is concave in the action and utility, and thus $\int_{\theta_l}^{\theta_h}\pi(\theta,\check{s}(\delta))h(\theta)d\theta$ is concave in δ. But then, by the previous strict inequality, it must be that, as desired,

$$\left.\frac{d}{d\delta}\int_{\theta_l}^{\theta_h}\pi(\theta,\check{s}(\delta))h(\theta)d\theta\right|_{\delta=0}>0.$$

Finally, let us see that $\kappa^1=0$, and $\kappa^N=1$. Note first that for $\theta\geq\theta_l^N$, $v^{-N}=v^{N-1}$. Thus, by definition $v^N\left(\theta_l^N\right)=v^{N-1}\left(\theta_l^N\right)$. But, by NEO, for all $\theta>\theta_l^N$, we have $\alpha^N-\alpha^{N-1}>0$ and thus $v_\theta^N>v_\theta^{N-1}$. Thus, $v^N\left(\bar{\theta}\right)>v^{N-1}\left(\bar{\theta}\right)$, and hence $z\left(\theta_l^N,\bar{\theta},\kappa^N\right)<0$, which by definition of $\tilde{\kappa}$ can only hold if $\kappa^N=H\left(\bar{\theta}\right)=1$. Similarly, $\kappa^1=0$. □

8.5 Proof of OB

Proposition 7 *Let s be Nash. Then,* (8) *and* (9) *hold.*

Proof Fix n. We will prove (9), with (8) analogous. We will consider perturbations that add or subtract workers in a continuous fashion immediately to the right or left of θ_h. We need to respect monotonicity and the integral condition, and make sure that our perturbed menus continue to hire an interval of workers (as opposed to a disconnected set thereof).[21]

If $\theta_l^{n+1}>\theta_h^n=\theta_h$, then (9) is automatic, since by Proposition 1 and the definition of PS, $\pi(\theta_h,\alpha,v)=0$ and $\alpha(\theta_h)=\alpha^{n+1}(\theta_h)$. So, assume $\theta_l^{n+1}=\theta_h$, and note that by Proposition 6, α is strictly increasing to the left of θ_h, and $a^{-n}=\alpha^{n+1}$ is strictly increasing to the right of θ_h.

Step 1 Let us first define a basic perturbation $(\hat{\alpha}(\cdot,y),\hat{v}(\cdot,y))$ indexed by y. Fix n and $0<\varepsilon<$

[21] This proof would be much easier if all crossing were strictly transverse. Then, we could use $\gamma(\cdot,\kappa)$ and vary κ holding fixed $v(\theta_l)$.

$\theta_h - \theta_l$. For y positive or negative, define

$$\hat{\alpha}(\theta, y) = \begin{cases} \alpha(\theta) & \text{if } \theta < \theta_h - \varepsilon \\ \max\{\alpha(\theta_h - \varepsilon), \min\{\alpha(\theta) + y, \alpha(\theta_h)\}\} & \text{if } \theta \geq \theta_h - \varepsilon \end{cases}.$$

That is, above $\theta_h - \varepsilon$, change actions by y but censor them to be at or above $\alpha(\theta_h - \varepsilon)$ and at or below $\alpha(\theta_h)$. Leave actions alone below $\theta_h - \varepsilon$. Note that monotonicity is preserved, and that $\hat{\alpha}$ is continuous at all (θ, y) where $\theta > \theta_h - \varepsilon$.

Define $\hat{v}(\theta, y) = v(\theta_l) + \int_{\theta_l}^{\theta} \hat{\alpha}(\tau, y) d\tau$. Because $\hat{\alpha}(\tau, y)$ is bounded and for each y, differentiable in y for almost all τ, with $\hat{\alpha}_y(\tau, y) \in \{0, 1\}$ wherever it is defined, $\hat{v}$ is continuously differentiable in (θ, y) wherever $\theta > \theta_h - \varepsilon$, with $\hat{v}_y(\theta_h, 0) = \varepsilon > 0$.

Step 2 Let us now use the basic perturbation to add or subtract workers near θ_h. Define $\hat{y}(\theta')$ implicitly by $\hat{v}(\theta', \hat{y}(\theta')) - v^{-n}(\theta') = 0$. Then $\hat{y}$ is well defined on an interval around θ_h, with

$$\hat{y}_{\theta'}(\theta') = \frac{a^{-n}(\theta') - \hat{\alpha}(\theta', \hat{y}(\theta'))}{\hat{v}_y(\theta', \hat{y}(\theta'))} \geq 0. \tag{20}$$

Further, when $\hat{y}(\theta') > 0$, then $\hat{v}(\theta, \hat{y}(\theta')) > v^{-n}(\theta)$ for all $\theta \in (\theta_l, \theta_h]$, and hence any crossing of zero by $\hat{v}(\cdot, \hat{y}(\theta')) - v^{-n}(\cdot)$ above θ_l occurs where $\theta > \theta_h$, and thus where

$$(\hat{v}(\theta, \hat{y}(\theta')) - v^{-n}(\theta))_\theta = \hat{\alpha}(\theta, \hat{y}(\theta')) - a^{-n}(\theta) = \alpha(\theta_h) - a^{-n}(\theta) < 0,$$

since $a^{-n}(\theta) > a^{-n}(\theta_h) \geq \alpha(\theta_h)$. Thus, indeed θ' is the unique crossing, and so $\varphi = 1$ for all $\theta \in (\theta_l, \theta')$, and $\varphi = 0$ outside of $[\theta_l, \theta']$. Similarly, if $\hat{y}(\theta') < 0$, then any crossing of zero by $\hat{v}(\cdot, \hat{y}(\theta')) - v^{-n}(\cdot)$ above θ_l occurs where $\theta < \theta_h$, and thus where $\hat{\alpha}(\theta, \hat{y}(\theta')) \leq \alpha(\theta)$, and hence

$$(\hat{v}(\theta, \hat{y}(\theta')) - v^{-n}(\theta))_\theta = \hat{\alpha}(\theta, \hat{y}(\theta')) - a^{-n}(\theta) \leq \alpha(\theta) - a^{-n}(\theta_h) < 0,$$

by NEO, and so again $\varphi = 1$ for all $\theta \in (\theta_l, \theta')$, and $\varphi = 0$ outside of $[\theta_l, \theta']$.

Step 3 Since this perturbation is feasible, it must be unprofitable. Let us show that this implies (9). To do so, let $j(\theta')$ be the profit from the perturbation. Then,

$$j(\theta') = \int_{\theta_l}^{\theta_h - \varepsilon} \pi(\theta, \alpha, v) h(\theta) d\theta + \int_{\theta_h - \varepsilon}^{\theta'} \pi(\theta, \hat{\alpha}(\cdot, \hat{y}(\theta')), \hat{v}(\cdot, \hat{y}(\theta')) h(\theta) d\theta,$$

since for $\theta < \theta_h - \varepsilon$, $\hat{\alpha} = \alpha$ and $\hat{v} = v$. Thus,

$$\begin{aligned} j_{\theta'}(\theta') &= \pi(\theta', \hat{\alpha}(\cdot, \hat{y}(\theta')), \hat{v}(\cdot, \hat{y}(\theta')) h(\theta') \\ &\quad + \hat{y}_\theta(\theta') \int_{\theta_h - \varepsilon}^{\theta'} \left(\pi_a(\theta, \hat{\alpha}(\cdot, \hat{y}(\theta')), \hat{v}(\cdot, \hat{y}(\theta')) \hat{\alpha}_y(\theta, \hat{y}(\theta')) - \hat{v}_y(\theta, \hat{y}(\theta')) \right) h(\theta) d\theta. \end{aligned}$$

To evaluate this at $\theta' = \theta_h$, note that $\hat{y}(\theta_h) = 0$, $\hat{\alpha}(\theta, 0) = \alpha(\theta)$, $\hat{\alpha}^n_y(\theta, 0) = 1$ for $\theta \in (\theta_h - \varepsilon, \theta_h)$, and zero outside of $[\theta_h - \varepsilon, \theta_h]$ and that $\hat{v}(\cdot, 0) = v$, and so, using (20) and $\hat{v}_y(\theta_h, 0) = \varepsilon$,

$$\begin{aligned} j_{\theta'}(\theta_h) &= \pi(\theta_h, \alpha, v)h(\theta_h) + \frac{a^{-n}(\theta_h) - \alpha(\theta_h)}{\varepsilon} \int_{\theta_h - \varepsilon}^{\theta_h} (\pi_a(\theta, \alpha, v) - \hat{v}_y(\theta, 0))\, h(\theta) d\theta \\ &= \pi(\theta_h, \alpha, v)h(\theta_h) + \left(a^{-n}(\theta_h) - \alpha(\theta_h)\right)(\pi_a(\tau, \alpha, v) - \hat{v}_y(\tau, 0))\, h(\tau) \end{aligned}$$

for some $\tau \in [\theta_h - \varepsilon, \theta_h]$ by the Mean Value Theorem, and where we note that $\hat{v}_y(\tau, 0) = \tau - (\theta_h - \varepsilon) \in [0, \varepsilon]$. But, for all $\varepsilon > 0$, this perturbation is feasible for all θ' in a neighborhood of θ_h, and so since (α, v) is optimal, we have $j_{\theta'}(\theta_h) = 0$. Taking $\varepsilon \to 0$, we have $\tau \to \theta_h$, and hence, canceling $h(\theta_h)$, we arrive at $0 = \pi(\theta_h, \alpha, v) + (a^{-n}(\theta_h) - \alpha(\theta_h))\, \pi_a(\theta_h, \alpha, v)$. Thus (9) holds, and we are done. □

8.6 Proofs for Section 5

Proof of Theorem 2 The first inequality in (11) follows from PP. Assume that $\pi^n(\theta, \alpha^n, v^n) > B^n(\alpha^n(\theta)) - \max_{n'} B^{n'}(\alpha^n(\theta))$. Then,

$$\begin{aligned} v^n(\theta) &= B^n(\alpha^n(\theta)) - c(\alpha^n(\theta), \theta) - \pi^n(\theta) \\ &< B^n(\alpha^n(\theta)) - c(\alpha^n(\theta), \theta) - (B^n(\alpha^n(\theta)) - \max_{n'} B^{n'}(\alpha^n(\theta))) \\ &= \max_{n'} B^{n'}(\alpha^n(\theta)) - c(\alpha^n(\theta), \theta), \end{aligned}$$

contradicting NP.

Let $\hat{n}$ be the firm that serves θ in an efficient allocation, let $\hat{\theta}$ be any type that $\hat{n}$ serves in equilibrium, and let $\hat{a} = \alpha^{\hat{n}}(\hat{\theta})$. By (11), $\hat{n}$ is the most efficient firm at action $\hat{a}$. Hence, $\hat{a} \in [a^{\hat{n}}_l, a^{\hat{n}}_h]$. Similarly, since $\hat{n}$ efficiently serves θ, we also have $\alpha^{\hat{n}}_*(\theta) \in [a^{\hat{n}}_l, a^{\hat{n}}_h]$. Thus, $\left|\hat{a} - \alpha^{\hat{n}}_*(\theta)\right| \leq d_2$.

Since θ can imitate $\hat{\theta}$, we have that

$$\begin{aligned} v^n(\theta) &> v^{\hat{n}}(\hat{\theta}) + c(\hat{a}, \hat{\theta}) - c(\hat{a}, \theta) \\ &= B^{\hat{n}}(\hat{a}) - c(\hat{a}, \hat{\theta}) - \pi^{\hat{n}}(\hat{\theta}) + c(\hat{a}, \hat{\theta}) - c(\hat{a}, \theta) \\ &= B^{\hat{n}}(\hat{a}) - \pi^{\hat{n}}(\hat{\theta}) - c(\hat{a}, \theta) \\ &\geq B^{\hat{n}}(\hat{a}) - c(\hat{a}, \theta) - d_1 \\ &= B^{\hat{n}}(\alpha^{\hat{n}}_*(\theta)) - c(\alpha^{\hat{n}}_*(\theta), \theta) - d_1 + (B^{\hat{n}}(\hat{a}) - c(\hat{a}, \theta) - (B^{\hat{n}}(\alpha^{\hat{n}}_*(\theta)) - c(\alpha^{\hat{n}}_*(\theta), \theta))) \\ &= v_*(\theta) - d_1 + (B^{\hat{n}}(\hat{a}) - c(\hat{a}, \theta) - (B^{\hat{n}}(\alpha^{\hat{n}}_*(\theta)) - c(\alpha^{\hat{n}}_*(\theta), \theta))) \\ &\geq v_*(\theta) - d_1 - \frac{1}{2} d_2^2 \delta, \end{aligned}$$

where the second inequality uses (11), and the third inequality follows from $(B^{\hat{n}}(\alpha^{\hat{n}}_*(\theta)) - c(\alpha^{\hat{n}}_*(\theta), \theta))_a =$

0, the definition of δ, and $\left|a_*^{\hat{n}}(\theta) - \hat{a}\right| < d_2$. Comparing the first and last terms gives the result.□

9 Appendix B: Proofs for Section 6

9.1 Proofs for Section 6.3

Remark 1 *For this and the next three subsections, we assume stacking, and whenever we fix n and s^{-n}, we assume s^{-n} satisfies $C1$ and $C2$.*

Proof of Lemma 3 That s^n is single dominant on some non-empty interval including θ^x is proven in the paragraph immediately following the statement of the Lemma. To see that OB implies NP, note that by $C1$ and stacking, $a^{-n} > \alpha_*$ for all θ above θ_h, and hence, using (10), the profits to poaching, $\pi(\cdot, a^{-n}, v^{-n})$, are falling everywhere above θ_h. (Similarly, poaching profits rise below θ_l.) It is thus enough to show that poaching just above θ_h does not make sense. This follows from (9), since

$$\begin{aligned} 0 &= \pi_a(\theta_h, \alpha, v)\left(a^{-n}(\theta_h) - \alpha(\theta_h)\right) + \pi(\theta_h, \alpha, v) \\ &> \pi(\theta_h, a^{-n}(\theta_h), v) - \pi(\theta_h, \alpha, v) + \pi(\theta_h, \alpha, v) \\ &= \pi(\theta_h, a^{-n}, v). \end{aligned}$$

where the inequality follows since π is strictly concave in a. □

9.2 Proofs for Section 6.4

We begin with a simple corollary to Lemma 2 that we will use repeatedly.

Corollary 5 *Assume that $\theta_l < \theta_h$ and $r(\theta_l, \theta_h) \geq 0$, and let $\tilde{s}(\theta_l, \theta_h) = (\tilde{\alpha}, \tilde{v})$. If $\tilde{\kappa}(\theta_l, \theta_h) = H(\theta_h)$, then $\pi(\theta_h, \tilde{\alpha}, \tilde{v}) > 0$, and if $\tilde{\kappa}(\theta_l, \theta_h) = H(\theta_l)$, then $\pi(\theta_l, \tilde{\alpha}, \tilde{v}) > 0$.*

This follows immediately from Lemma 2. If $\tilde{\kappa}(\theta_l, \theta_h) = H(\theta_h)$, and $\pi(\theta_h, \tilde{\alpha}, \tilde{v}) \leq 0$ then the the integrand in the objective function (4) is strictly negative everywhere on (θ_l, θ_h), contradicting $r(\theta_l, \theta_h) \geq 0$, and similarly if $\tilde{\kappa}(\theta_l, \theta_h) = H(\theta_l)$, and $\pi(\theta_l, \tilde{\alpha}, \tilde{v}) \leq 0$.

Proof of Lemma 4 Note that r is continuous, since $\tilde{\kappa}$ is continuous in (θ_l, θ_h), γ is continuous in κ, $\tilde{v}$ is continuous in $(\theta_l, \theta_h, \kappa)$, and the integral in (4) is continuous in its endpoints. Since the set $\{\theta_l, \theta_h | \underline{\theta} \leq \theta_l \leq \theta_h \leq \bar{\theta}\}$ is compact, r has a maximum. Part (i) follows since $\tilde{\kappa} \in [0,1]$ using Lemma 6, and hence $\tilde{\alpha}$ is monotone, and since $\tilde{v}(\theta) = \tilde{v}(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \tilde{\alpha} d\tau$ by construction. To see (ii), consider any maximizer (θ_l, θ_h) of r at which $\tilde{v}(\theta_h) > v^{-n}(\theta_h)$. Then, for all θ_h' on a neighborhood of θ_h, (a) $\tilde{\kappa}(\theta_l, \theta_h') = H(\theta_h')$, (b) $\tilde{\kappa}$ is differentiable in its second argument and surplus at θ_l remains fixed at $v^{-n}(\theta_l)$, and (c) $\tilde{s}(\theta_l, \theta_h')$ is feasible in $\mathcal{P}(\theta_l, \theta_h)$. But then, since $\tilde{s}(\theta_l, \theta_h)$ was optimal in $\mathcal{P}(\theta_l, \theta_h)$, we have that $\int_{\theta_l}^{\theta_h} (\pi(\theta, \tilde{s}(\theta_l, \theta_h)))_{\theta_h} h(\theta) d\theta$ is well-defined and equal to 0. Hence,

$$\begin{aligned} r_{\theta_h}(\theta_l, \theta_h) &= \left(\int_{\theta_l}^{\theta_h} \pi(\theta, \tilde{s}(\theta_l, \theta_h)) h(\theta) d\theta \right)_{\theta_h} \\ &= \pi(\theta_h, \tilde{s}(\theta_l, \theta_h)) h(\theta) + \int_{\theta_l}^{\theta_h} (\pi(\theta, \tilde{s}(\theta_l, \theta_h)))_{\theta_h} h(\theta) d\theta \\ &= \pi(\theta_h, \tilde{s}(\theta_l, \theta_h)) h(\theta). \end{aligned}$$

But, since $\tilde{\kappa}(\theta_l, \theta_h) = H(\theta_h)$, and since (θ_l, θ_h) is a maximum of r, and so $r(\theta_l, \theta_h) > 0$ by $C2$, we have $\pi(\theta_h, \tilde{s}(\theta_l, \theta_h)) > 0$ by Corollary 5, and thus $r_{\theta_h}(\theta_l, \theta_h) > 0$. Since (θ_l, θ_h) is optimal, it must thus be that $\theta_h = \bar{\theta}$. Similarly, if $\tilde{v}(\theta_l) > v^{-n}(\theta_l)$, then $\theta_l = \underline{\theta}$. But then, in all cases, $\tilde{s}(\theta_l, \theta_h)$ is single dominant on (θ_l, θ_h), using stacking and $C1$. Part (iii) follows immediately, with the equality of payoffs following as the relevant domains of integration agree. □

Lemma 8 *There exist $\underline{m}$ and $\overline{m}$ with $\underline{m} \leq \theta^x \leq \overline{m}$ such that $\pi(\theta, a^{-n}, v^{-n})$ is strictly positive if $\theta \in (\underline{m}, \overline{m})$, strictly negative and strictly increasing if $\theta < \underline{m}$, and strictly negative and strictly decreasing if $\theta > \overline{m}$.*

Proof By stacking and $C1$, for $\theta > \theta^x$,

$$a^{-n}(\theta) > \gamma(\theta, 0) \geq \gamma(\theta, H(\theta)) = \alpha_*(\theta),$$

and so $\pi_a(\theta, a^{-n}, v^{-n}) < 0$. Hence, anywhere that a^{-n} is differentiable, we have by (10) that $(\pi(\theta, a^{-n}, v^{-n}))_\theta < 0$. Further, at any point where a^{-n} jumps, say from a_l to a_h, we have, since v^{-n} is continuous, and since $a_h > a_l > \alpha_*(\theta)$ that $\pi(\theta, a_h, v^{-n}) - \pi(\theta, a_l, v^{-n}) < 0$. Hence $\pi(\cdot, a^{-n}, v^{-n})$ is strictly decreasing on $[\theta^x, \bar{\theta}]$, and so single-crosses 0 from above at most once on $[\theta^x, \bar{\theta}]$. If such a crossing exists, define $\overline{m}$ as the crossing. If $\pi(\bar{\theta}, a^{-n}, v^{-n}) > 0$, take $\overline{m} = \bar{\theta}$, and if $\pi(\theta^x, a^{-n}, v^{-n}) < 0$, take $\bar{m} = \theta^x$. Construct $\underline{m}$ similarly. □

Lemma 9 *Let (α, v) be any feasible menu for n with $\pi(\theta, \alpha, v) \geq 0$ everywhere, and let v be dominant on (τ_l, τ_h). Then, $(\tau_l, \tau_h) \cap [\underline{m}, \overline{m}] \neq \varnothing$.*

Proof Assume $\tau_l \geq \overline{m} \geq \theta^x$. Then, $v(\tau_l) = v^{-n}(\tau_l)$ by definition of dominance and since v and v^{-n} are continuous. Since for all $\theta \in (\tau_l, \tau_h)$

$$v(\tau_l) + \int_{\tau_l}^{\theta} \alpha(\tau) d\tau = v(\theta) > v^{-n}(\theta) = v^{-n}(\tau_l) + \int_{\tau_l}^{\theta} a^{-n}(\tau) d\tau$$

it follows that there is $\tau \in (\tau_l, \tau_h)$ where $\alpha(\tau) > a^{-n}(\tau)$. But, since $\tau > \overline{m} \geq \theta^x$, and using $C1$, it follows that $a^{-n}(\tau) > \alpha_*(\tau)$, and so

$$\pi(\tau, \alpha(\tau), v(\tau)) < \pi(\tau, \alpha^{-n}(\tau), v(\tau)) < \pi(\tau, \alpha^{-n}(\tau), v^{-n}(\tau)) < 0,$$

a contradiction. Similarly, it cannot be that $\tau_h \leq \underline{m}$. □

Proof of Proposition 2 Using Proposition 3, we can wlog assume that (α, v) loses money nowhere. Recall that

$$\Pi(s) = \int_{\underline{\theta}}^{\bar{\theta}} \pi(\theta, \alpha, v)\varphi(\theta, s)h(\theta)d\theta. \tag{21}$$

Assume that v dominates v^{-n} on an interval I_H with $\theta^x \leq \underline{I}_H \leq \overline{m} \leq \bar{I}_H$. In this case, define $\overline{m}^* = \bar{I}_H$. If there is no such interval, define $\overline{m}^* = \overline{m}$. Similarly, if v dominates v^{-n} on an interval I_L with $\underline{I}_L \leq \underline{m} \leq \bar{I}_L \leq \theta^x$, then define $\underline{m}^* = \underline{I}_L$, and if there is no such interval, define $\underline{m}^* = \underline{m}$.

Consider first any positive lengthed interval $J \subseteq [\overline{m}^*, \bar{\theta}]$ on which $v = v^{-n}$, and such that $\int_J \varphi(\theta, s)d\theta > 0$. Then, $\alpha = a^{-n}$ on this interval, and so, since $\overline{m}^* \geq \overline{m}$, $\pi(\theta, \alpha, v) < 0$ for all $\theta > \overline{m}^*$. Hence, excluding J from the domain of the integral in (21) increases the integral in (21).

By Lemma 9, there is no positive lengthed interval $J = (\underline{J}, \bar{J})$ with $\underline{J} \geq \overline{m}^*$ or and $\bar{J} < \underline{m}^*$ on which v is dominant. We thus have

$$\Pi(s) \leq \int_{\underline{m}^*}^{\overline{m}^*} \pi(\theta, \alpha, v)\varphi(\theta, s)h(\theta)d\theta. \tag{22}$$

Define $\hat{v} = \max(v, v^{-n})$, with associated $\hat{\alpha}$, where at all θ where $v(\theta) \geq v^{-n}(\theta)$, we can take $\hat{\alpha} = \alpha$, and at almost all θ where $v(\theta) \leq v^{-n}(\theta)$, we can take $\hat{\alpha} = a^{-n}$ (on any interval where $v(\theta) = v^{-n}(\theta)$, $\alpha = a^{-n}$ almost everywhere, and so there is a zero measure set where the two definitions might be in conflict). But then, everywhere that $\varphi(\theta, s)$ is positive (and so $v(\theta) \geq v^{-n}(\theta)$), we have $\pi(\theta, \hat{\alpha}, \hat{v}) = \pi(\theta, \alpha, v)$, and so, from (22),

$$\Pi(s) \leq \int_{\underline{m}^*}^{\bar{m}^*} \pi(\theta, \hat{\alpha}, \hat{v})\, \varphi(\theta, s)\, h(\theta)\, d\theta.$$

Consider any $\theta \in (\underline{m}^*, \overline{m}^*)$ such that $\varphi(\theta, s) < 1$. Then, $v(\theta) \leq v^{-n}(\theta)$, and so $\hat{v}(\theta) = v^{-n}(\theta)$, and $\hat{\alpha}(\theta) = a^{-n}(\theta)$ almost everywhere. And, since by construction, φ is 1 on I_H and I_L (if these sets exist), it follows that $\theta \in [\underline{m}, \overline{m}]$, and so $\pi(\theta, \hat{\alpha}, \hat{v}) = \pi(\theta, a^{-n}, v^{-n}) \geq 0$. We thus have

$$\Pi(s) \leq \int_{\underline{m}^*}^{\overline{m}^*} \pi(\theta, \hat{\alpha}, \hat{v})h(\theta)d\theta.$$

But, $\hat{v} \geq v^{-n}$ by construction, and so (5) and (6) are satisfied in $\mathcal{P}(\underline{m}^*, \overline{m}^*)$, while $\hat{\alpha}$ was chosen to be a subgradient of the convex function $\max(v, v^{-n})$, and hence (7) holds as well. Thus, $(\hat{\alpha}, \hat{v})$ is feasible in the relaxed problem $\mathcal{P}(\underline{m}^*, \bar{m}^*)$, from which $\Pi(s) \leq r(\underline{m}^*, \overline{m}^*)$. □

9.3 Proofs for Section 6.5

In this section, we establish that the firm has an essentially unique best response. We begin with local properties of r and then use those properties to show that r has a unique maximum.

9.3.1 Local Properties of r

We first study the properties of r, including its strict local concavity properties. Write f_x^+ and f_x^- for the right and left derivatives of f with respect to x.

Remark 2 *Fix a maximal rectangle $\tilde{R} = [\iota_l, \iota_h] \times [\iota_l', \iota_h']$ as defined in Section 6.5. On $[\iota_l, \iota_h]$ and $[\iota_l', \iota_h']$, we can take α^{-n} to be continuous by C1, and hence v^{-n} to be continuously differentiable.*

Lemma 10 *Considered as a function on $\tilde{R}$, r is continuously differentiable, with*

$$r_{\theta_h}(\theta_l, \theta_h) = \left(\pi_a(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v})(a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa})) + \pi(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v})\right) h(\theta_h), \tag{23}$$

and

$$r_{\theta_l}(\theta_l, \theta_h) = \left(\pi_a\left(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \tilde{v}\right)\left(\gamma(\theta_l, \tilde{\kappa}) - a^{-n}(\theta_l)\right) - \pi(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \tilde{v})\right) h(\theta_l). \tag{24}$$

Proof The right side of (23) has the same form as (9). As in the analysis of OB in Section 8.5, this is the value of increasing θ_h by increasing effort immediately to the left of θ_h, and, since $\gamma\left(\cdot, \tilde{\kappa}\left(\theta_l, \theta_h\right)\right)$ solves the relaxed problem, this perturbation is as good as anything. Alternatively, differentiate r, and manipulate, using integration by parts and (19). The proof of (24) is similar. On $\tilde{R}$, all the terms of r_{θ_h} and r_{θ_l} are continuous. Hence, r is continuously differentiable. □

As a coherence check, along the lower boundary of $\tilde{R}$,

$$r_{\theta_h}^+(\theta_l, \theta_h) = \lim_{\varepsilon \downarrow 0} \frac{r(\theta_l, \theta_h + \varepsilon) - r(\theta_l, \theta_h)}{\varepsilon} = \lim_{\varepsilon \downarrow 0} r_{\theta_h}(\theta_l, \theta_h + \varepsilon) = \left(r|_{\tilde{R}}\right)_{\theta_h}(\theta_l, \theta_h), \tag{25}$$

where the second equality uses L'Hôpital's rule and continuity of r_{θ_h} on (ι_l', ι_h'). Things are similar on the other boundaries of $\tilde{R}$.

Recall that Θ is the subset of R on which $z(\theta_l, \theta_h, \tilde{\kappa}(\theta_l, \theta_h)) = 0$. Where there is no ambiguity, we will write $\tilde{\kappa}$ for $\tilde{\kappa}\left(\theta_l, \theta_h\right)$.

Lemma 11 *On $\tilde{R} \cap \Theta$, we have*

$$\left(\gamma(\theta_l, \tilde{\kappa})\right)_{\theta_l} > \gamma_\kappa(\theta_l, \tilde{\kappa})\tilde{\kappa}_{\theta_l} \geq 0 \text{ and } \left(\gamma(\theta_h, \tilde{\kappa})\right)_{\theta_h} > \gamma_\kappa(\theta_h, \tilde{\kappa})\tilde{\kappa}_{\theta_h} \geq 0,$$

with

$$\begin{vmatrix} \left(\gamma(\theta_l, \tilde{\kappa})\right)_{\theta_l} & \left(\gamma(\theta_l, \tilde{\kappa})\right)_{\theta_h} \\ \left(\gamma(\theta_h, \tilde{\kappa})\right)_{\theta_l} & \left(\gamma(\theta_h, \tilde{\kappa})\right)_{\theta_h} \end{vmatrix} > 0. \tag{26}$$

Proof Note that

$$(\gamma(\theta_l,\tilde{\kappa}))_{\theta_l} = \gamma_\theta(\theta_l,\tilde{\kappa}) + \gamma_\kappa(\theta_l,\tilde{\kappa})\tilde{\kappa}_{\theta_l} > \gamma_\kappa(\theta_l,\tilde{\kappa})\tilde{\kappa}_{\theta_l},$$

since $\gamma_\theta > 0$ using that $\tilde{\kappa} \in [0,1]$. But, since $z(\theta_l,\theta_h,\tilde{\kappa}) = 0$, we have $\tilde{\kappa}_{\theta_l} = -z_{\theta_l}/z_\kappa < 0$ using the discussion around (13). Intuitively, when θ_l is increased, v must become steeper, and this is accomplished via higher actions and hence a lower κ. Thus $\gamma_\kappa(\theta_l,\tilde{\kappa})\tilde{\kappa}_{\theta_l} > 0$, since $\gamma_\kappa < 0$. Similarly, $\tilde{\kappa}_{\theta_h} < 0$, and so $(\gamma(\theta_h,\tilde{\kappa}))_{\theta_h} > \gamma_\kappa(\theta_h,\tilde{\kappa})\tilde{\kappa}_{\theta_h} > 0$.

To see (26), note that

$$\begin{vmatrix} (\gamma(\theta_l,\tilde{\kappa}))_{\theta_l} & (\gamma(\theta_l,\tilde{\kappa}))_{\theta_h} \\ (\gamma(\theta_h,\tilde{\kappa}))_{\theta_l} & (\gamma(\theta_h,\tilde{\kappa}))_{\theta_h} \end{vmatrix} = \begin{vmatrix} \gamma_\theta(\theta_l,\tilde{\kappa}) + \gamma_\kappa(\theta_l,\tilde{\kappa})\tilde{\kappa}_{\theta_l} & \gamma_\kappa(\theta_l,\tilde{\kappa})\tilde{\kappa}_{\theta_h} \\ \gamma_\kappa(\theta_h,\tilde{\kappa})\tilde{\kappa}_{\theta_l} & \gamma_\theta(\theta_h,\tilde{\kappa}) + \gamma_\kappa(\theta_h,\tilde{\kappa})\tilde{\kappa}_{\theta_h} \end{vmatrix}$$

$$> \begin{vmatrix} \gamma_\kappa(\theta_l,\tilde{\kappa})\tilde{\kappa}_{\theta_l} & \gamma_\kappa(\theta_l,\tilde{\kappa})\tilde{\kappa}_{\theta_h} \\ \gamma_\kappa(\theta_h,\tilde{\kappa})\tilde{\kappa}_{\theta_l} & \gamma_\kappa(\theta_h,\tilde{\kappa})\tilde{\kappa}_{\theta_h} \end{vmatrix} = 0,$$

since each of the four terms on the main diagonal in the second expression is strictly positive. □

Lemma 12 *Consider r as a function on $\tilde{R} \cap \Theta$. Then, $r_{\theta_l\theta_h} < 0$. If $r_{\theta_h}(\theta_l,\theta_h) = 0$, then $r_{\theta_h\theta_h}(\theta_l,\theta_h) < 0$, if $r_{\theta_l}(\theta_l,\theta_h) = 0$, then $r_{\theta_l\theta_l}(\theta_l,\theta_h) < 0$, and if $r_{\theta_l}(\theta_l,\theta_h) = r_{\theta_h}(\theta_l,\theta_h) = 0$, then r is locally strictly concave at (θ_l,θ_h).*

Proof From (23),

$$\begin{aligned} \frac{r_{\theta_h\theta_l}(\theta_l,\theta_h)}{h(\theta_h)} &= \pi_{aa}(\theta_h,\gamma(\theta_h,\tilde{\kappa}),\tilde{v})\,(\gamma(\theta_h,\tilde{\kappa}))_{\theta_l}\left(a^{-n}(\theta_h) - \gamma(\theta_h,\tilde{\kappa})\right) \\ &\quad + \pi_a(\theta_h,\gamma(\theta_h,\tilde{\kappa}),\tilde{v})(-(\gamma(\theta_h,\tilde{\kappa}))_{\theta_l}) \\ &\quad + \pi_a(\theta_h,\gamma(\theta_h,\tilde{\kappa}),\tilde{v})\,(\gamma(\theta_h,\tilde{\kappa}))_{\theta_l}, \\ &= \pi_{aa}(\theta_h,\gamma(\theta_h,\tilde{\kappa}),\tilde{v})\,(\gamma(\theta_h,\tilde{\kappa}))_{\theta_l}\left(a^{-n}(\theta_h) - \gamma(\theta_h,\tilde{\kappa})\right) \end{aligned} \tag{27}$$

and similarly, from (24),

$$\frac{r_{\theta_l\theta_h}(\theta_l,\theta_h)}{h(\theta_l)} = \pi_{aa}(\theta_l,\gamma(\theta_l,\tilde{\kappa}),\tilde{v})\,(\gamma^n(\theta_l,\tilde{\kappa}))_{\theta_h}\left(\gamma(\theta_l,\tilde{\kappa}) - a^{-n}(\theta_l)\right).\text{[22]} \tag{28}$$

To see that $r_{\theta_l\theta_h} < 0$, start from (28), and note that $\pi_{aa} < 0$, that $(\gamma^n(\theta_l,\tilde{\kappa}))_{\theta_h} > 0$, and that $\gamma(\theta_l,\tilde{\kappa}) - a^{-n}(\theta_l) > 0$.

Note next that since $z(\theta_l,\theta_h,\tilde{\kappa}) = 0$, $\tilde{v}(\theta_l) = v^{-n}(\theta_l)$ and $\tilde{v}(\theta_h) = v^{-n}(\theta_h)$. Note that $\pi_a(\theta_h,\gamma(\cdot,\tilde{\kappa}),\tilde{v}) =_s \tilde{\kappa} - H(\theta_h) \leq 0$,since $\tilde{\kappa} \in [H(\theta_l), H(\theta_h)]$. Similarly, $\pi_a(\theta_l,\gamma(\cdot,\tilde{\kappa}),\tilde{v}) \geq 0$.

[22] These two expressions must of course be equal, but it is convenient to express them in these two different ways.

Assume that $r_{\theta_h}(\theta_l, \theta_h) = 0$. Then using (23),

$$
\begin{aligned}
\frac{r_{\theta_h\theta_h}(\theta_l,\theta_h)}{h(\theta_h)} &= \Big(1 + \pi_{aa}(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v})\,(\gamma(\theta_h, \tilde{\kappa}))_{\theta_h}\Big)\left(a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa})\right) \\
&\quad + \pi_a(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v})(a_\theta^{-n}(\theta_h) - (\gamma(\theta_h, \tilde{\kappa}))_{\theta_h}) \\
&\quad + \gamma(\theta_h, \tilde{\kappa}) + \pi_a(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v})\,(\gamma(\theta_h, \tilde{\kappa}))_{\theta_h} - a^{-n}(\theta_h)
\end{aligned}
$$

where the term involving h_θ disappears since $r_{\theta_h} = 0$, and where we use that $\tilde{v}(\theta_h) = v^{-n}(\theta_h)$, and hence $(\tilde{v}(\theta_h))_{\theta_h} = (v^{-n}(\theta_h))_{\theta_h} = a^{-n}(\theta_h)$. Cancelling,

$$
\begin{aligned}
\frac{r_{\theta_h\theta_h}(\theta_l,\theta_h)}{h(\theta_h)} &= \pi_{aa}(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v})\,(\gamma(\theta_h, \tilde{\kappa}))_{\theta_h}\left(a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa})\right) \\
&\quad + \pi_a(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v}) a_\theta^{-n}(\theta_h) \\
&\leq \pi_{aa}(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v})\,(\gamma(\theta_h, \tilde{\kappa}))_{\theta_h}\left(a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa})\right) \\
&< 0,
\end{aligned}
\tag{29}
$$

where the first inequality uses that $\pi_a(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v}) \leq 0$ and the second uses that $\pi_{aa} < 0$, that by Lemma 11, $(\gamma(\theta_h, \tilde{\kappa}))_{\theta_h} > 0$, and that by stacking, $C1$, and $\tilde{\kappa} \in [0, 1]$, $a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa}) > 0$.[23]

Similarly, taking cancellations as before, if $r_{\theta_l} = 0$, then

$$
\begin{aligned}
\frac{r_{\theta_l\theta_l}(\theta_l,\theta_h)}{h(\theta_l)} &= \pi_{aa}(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \tilde{v})\,(\gamma(\theta_l, \tilde{\kappa}))_{\theta_l}\left(\gamma(\theta_l, \tilde{\kappa}) - a^{-n}(\theta_l)\right) \\
&\quad - \pi_a(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \tilde{v}) a_\theta^{-n}(\theta_l) \\
&\leq \pi_{aa}(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \tilde{v})\,(\gamma(\theta_l, \tilde{\kappa}))_{\theta_l}\left(\gamma(\theta_l, \tilde{\kappa}) - a^{-n}(\theta_l)\right) \\
&< 0.
\end{aligned}
\tag{30}
$$

For strict local concavity, it remains to show that where $r_{\theta_l} = r_{\theta_h} = 0$, we have $d \equiv r_{\theta_l\theta_l} r_{\theta_h\theta_h} - r^2_{\theta_l\theta_h} > 0$. From (29) and (30),

$$
\begin{aligned}
\frac{r_{\theta_l\theta_l} r_{\theta_h\theta_h}}{h(\theta_l)h(\theta_h)} &\geq \pi_{aa}(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v})\,(\gamma(\theta_h, \tilde{\kappa}))_{\theta_h}\left(a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa})\right) \\
&\quad \times \pi_{aa}(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \tilde{v})\,(\gamma(\theta_l, \tilde{\kappa}))_{\theta_l}\left(\gamma(\theta_l, \tilde{\kappa}) - a^{-n}(\theta_l)\right),
\end{aligned}
$$

[23] To be careful, a_θ^{-n}, and hence $r_{\theta_h\theta_h}$ may not be everywhere defined, in particular, where the opponent changes. But, since a^{-n} is increasing, $\liminf_{\varepsilon\downarrow 0} a_\theta^{-n}(\theta_h + \varepsilon) \geq 0$ and $\liminf_{\varepsilon\downarrow 0} a_\theta^{-n}(\theta_h - \varepsilon) \geq 0$, and so, since $\pi_a(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v}) \leq 0$, we have $\limsup_{\varepsilon\downarrow 0} r_{\theta_h\theta_h}(\theta_l, \theta_h + \varepsilon) < 0$, and $\limsup_{\varepsilon\downarrow 0} r_{\theta_h\theta_h}(\theta_l, \theta_h - \varepsilon) < 0$. We henceforth ignore this technical detail.

while from (27) and (28),

$$\begin{aligned}\frac{r_{\theta_l\theta_h} r_{\theta_h\theta_l}}{h(\theta_l)h(\theta_h)} &= \pi_{aa}(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \tilde{v}) \left(\gamma(\theta_l, \tilde{\kappa})\right)_{\theta_h} \left(\gamma(\theta_l, \tilde{\kappa}) - a^{-n}(\theta_l)\right) \\ &\quad \times \pi_{aa}(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v}) \left(\gamma(\theta_h, \tilde{\kappa})\right)_{\theta_l} \left(a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa})\right).\end{aligned}$$

Collecting the three positive terms $h(\theta_l)h(\theta_h)$, $(\gamma(\theta_l, \tilde{\kappa}) - a^{-n}(\theta_l))\,(a^{-n}(\theta_h) - \gamma(\theta_h, \tilde{\kappa}))$, and $\pi_{aa}(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \tilde{v})\pi_{aa}(\theta_h, \gamma(\theta_h, \tilde{\kappa}), \tilde{v})$, it suffices that

$$\left(\gamma(\theta_h, \tilde{\kappa})\right)_{\theta_h} \left(\gamma(\theta_l, \tilde{\kappa})\right)_{\theta_l} - \left(\gamma(\theta_l, \tilde{\kappa})\right)_{\theta_h} \left(\gamma(\theta_h, \tilde{\kappa})\right)_{\theta_l} > 0,$$

which follows from Lemma 11. □

9.3.2 Essentially Unique Optimality

Say that $f : \mathbb{R} \to \mathbb{R}$ has a critical point at x if $f_x^-(x) f_x^+(x) \leq 0$, so that f_x at least weakly changes sign at x. This includes the case where f is differentiable at x and $f_x(x) = 0$. Recall that K is the set of kink points of v^{-n}. By $C1$, $|K| \leq N - 1$.

Lemma 13 *Any maximum of r is in $R = [\underline{\theta}, \theta^x] \times [\theta^x, \bar{\theta}]$.*

Proof Consider any (θ_l, θ_h) with $\theta_h < \theta^x$, and assume (θ_l, θ_h) is a maximum of r. Then, $\tilde{v}(\theta_h) = v^{-n}(\theta_h)$ by Lemma 4, and so, since $\tilde{\kappa} < 1$, it follows from stacking and the definition of θ^x that $a^{-n} < \gamma(\cdot, \tilde{\kappa})$ for $\theta < \theta^x$, and so v crosses v^{-n} from below at θ_h, contradicting the definition of θ_h. Thus, $\theta_h \geq \theta^x$. Similarly, $\theta_l \leq \theta^x$. □

Next, we collect some basic properties of z.

Lemma 14 *For each $\kappa \in [0, 1]$, $z(\cdot, \cdot, \kappa)$ is strictly increasing on R, with $z(\theta_l, \theta^x, \kappa) < 0$ for all $\theta_l < \theta^x$.*

Proof Recall from the discussion around (13) that on each $\tilde{R}$, $z_{\theta_h}(\theta_l, \theta_h, \kappa) = a^{-n}(\theta_h) - \gamma(\theta_h, \kappa) =_s \theta_h - \theta^x$, and so $z(\theta_l, \cdot, \kappa)$ is strictly single-troughed with minimum at θ^x. But, $z(\theta_l, \theta_l, \kappa) = 0$, and hence $z(\theta_l, \theta^x, \kappa) < 0$ for all $\theta_l < \theta^x$. Similarly $z_{\theta_l}(\theta_l, \theta_h, \kappa) = -a^{-n}(\theta_l) + \gamma(\theta_l, \kappa) > 0$ for $\theta_l < \theta^x$. □

Recall that the locus L_N is defined by $z(\theta_l, \theta_h, H(\theta_l)) = 0$, and the locus L_S is defined by $z(\theta_l, \theta_h, H(\theta_h)) = 0$. Note that z is differentiable on each $\tilde{R}$. Hence, since z_{θ_l}, z_{θ_h}, and z_κ are strictly positive, L_N is continuous, strictly decreasing, and by definition of z, goes through (θ^x, θ^x). The locus L_S, which lies below L_N, has the same properties.

Let us show that on or below L_S, r is strictly increasing in θ_h where r is positive, and on or above L_N, r is strictly decreasing in θ_l where r is positive.

Lemma 15 *On or below L_S, $r_{\theta_h}(\theta_l,\theta_h)$ exists and equals $\pi(\theta_h, \tilde{s}(\theta_l,\theta_h))h(\theta_h)$, and if $r(\theta_l,\theta_h)>0$, then $r_{\theta_h}(\theta_l,\theta_h)>0$. On or above L_N, $r_{\theta_l}(\theta_l,\theta_h)$ exists and equals $-\pi(\theta_l,\gamma(\cdot,\tilde{s}(\theta_l,\theta_h))h(\theta_l)$, and if $r(\theta_l,\theta_h)>0$, then $r_{\theta_l}(\theta_l,\theta_h)<0$.*

Proof Fix (θ_l,θ_h) below L_S. Then, $z(\theta_l,\theta_h,H(\theta_h))<0$, and so by definition, $\tilde{\kappa}(\theta_l,\theta_h)=H(\theta_h)$, and by Proposition 1, $\tilde{v}(\theta_l)=v^{-n}(\theta_l)$, and thus $\tilde{v}(\theta_h)>v^{-n}(\theta_h)$. But then, for each ε small, the menu $\tilde{s}(\theta_l,\theta_h+\varepsilon)$ is feasible in $\mathcal{P}(\theta_l,\theta_h)$ and so by the Envelope Theorem (or by manipulation involving Lemma 7), $r_{\theta_h}(\theta_l,\theta_h)=\pi(\theta_h,\tilde{s}(\theta_l,\theta_h))h(\theta_h)$. If $r(\theta_l,\theta_h)>0$, then, since $\tilde{\kappa}=H(\theta_h)$, we have by Corollary 5 that $\pi(\theta_h,\tilde{s}(\theta_l,\theta_h))>0$, and hence $r_{\theta_h}(\theta_l,\theta_h)>0$.

Consider next $(\theta_l,\theta_h)\in L_S$. Since for each $\varepsilon>0$, $(\theta_l,\theta_h-\varepsilon)$ is below L_S, $r_{\theta_h}(\theta_l,\theta_h-\varepsilon)=\pi(\theta_h-\varepsilon,\tilde{s}(\theta_l,\theta_h-\varepsilon))h(\theta_h-\varepsilon)$ by the previous step. It thus follows as in (25) that $r^-_{\theta_h}(\theta_l,\theta_h)=\pi(\theta_h,\tilde{s}(\theta_l,\theta_h))h(\theta_h)$, where we note that on L_S, $\tilde{v}(\theta_h)=v^{-n}(\theta_h)$. Finally, from (23) and the discussion immediately following Remark 2, and again exploiting that above L_S, $\tilde{v}(\theta_h)=v^{-n}(\theta_h)$,

$$
\begin{aligned}
r^+_{\theta_h}(\theta_l,\theta_h) &= \lim_{\varepsilon\downarrow 0} r_{\theta_h}(\theta_l,\theta_h+\varepsilon)\\
&= \lim_{\varepsilon\downarrow 0}\left(\pi_a(\theta_h+\varepsilon,\tilde{s}(\theta_l,\theta_h+\varepsilon)),v^{-n}\right)\left(a^{-n}(\theta_h+\varepsilon)-\gamma(\theta_h+\varepsilon,H(\theta_h+\varepsilon))\right)\Big) h(\theta_h+\varepsilon)\\
&\quad +\lim_{\varepsilon\downarrow 0}\left(\pi(\theta_h+\varepsilon,\tilde{s}(\theta_l,\theta_h+\varepsilon)),v^{-n}\right)\Big) h(\theta_h+\varepsilon)\\
&= \pi(\theta_h,\tilde{s}(\theta_l,\theta_h))h(\theta_h).
\end{aligned}
$$

This follows since $a^{-n}(\cdot)-\gamma(\cdot,H(\theta_h))$ is bounded and since $\lim_{\varepsilon\downarrow 0}\pi_a(\theta_h+\varepsilon,\tilde{s}(\theta_l,\theta_h+\varepsilon))=0$ using that γ and v^{-n} are continuous and that on L_S, $\tilde{\kappa}=H(\theta_h)$, and hence $\pi_a(\theta_h,\tilde{s}(\theta_l,\theta_h))=0$ by definition of γ. But then, $r^+_{\theta_h}(\theta_l,\theta_h)=r^-_{\theta_h}(\theta_l,\theta_h)$, and so $r_{\theta_h}(\theta_l,\theta_h)$ exists and has the claimed value. The proof for (θ_l,θ_h) above L_N is similar. □

Now let us formalize the assumption that L_S hits the western boundary of R.

Assumption 1 $z\left(\underline{\theta},\bar{\theta},1\right)\geq 0$.

Note in particular that since z is increasing, Assumption 1 implies that $z\left(\theta_l,\bar{\theta},1\right)>0$ for all $\theta_l>\underline{\theta}$, so that L_S does not intersect with the northern boundary of R.

Define θ_T by $z(\underline{\theta},\theta_T,0)=0$ if there is such a $\theta_T\geq\theta^x$, and by $\theta_T=\bar{\theta}$ otherwise. This is the latitude at which L_N exits R. For Firm 1, $\theta^x=\underline{\theta}$, and hence $\theta_T=\underline{\theta}$. Let $A=\{(\underline{\theta},\theta_h)|\theta_h\geq\theta_T\}$. By Lemma 15, any maximum of r occurs in $\Theta\cup A$. In particular, if r is strictly positive (as it must be at an optimum), then below L_S, $r_{\theta_h}>0$, contradicting optimality, and above L_N, $r_{\theta_l}<0$, contradicting optimality unless $\theta_l=\underline{\theta}$. Thus, recalling that $\Theta(\theta_h)$ is the interval of θ_l such that $(\theta_l,\theta_h)\in\Theta\cup A$, and that $\psi(\theta_h)=\max_{\theta_l\in\Theta(\theta_h)}r(\theta_l,\theta_h)$, we have that

$$\max_{\{(\theta_l,\theta_h)|\theta_h\geq\theta_l\}} r(\theta_l,\theta_h)=\max_{\Theta\cup A} r(\theta_l,\theta_h)=\max_{\theta_h}\psi(\theta_h).$$

Given this construction, let us begin by considering the maximization problem as one moves east to west. Recall that D is the set of θ_h such that ψ is strictly positive.

Lemma 16 *Fix $\theta_h \in D$. Then on $\Theta(\theta_h)$, $r(\cdot, \theta_h)$ is strictly single-peaked and has a unique maximum $\lambda(\theta_h)$.*

Proof This is trivial for $\theta_h > \theta_T$, since $\Theta(\theta_h) = \{\underline{\theta}\}$. Fix $\theta_h \leq \theta_T$. Let the (closed) interval $\Theta(\theta_h)$ be denoted $[\tau_l, \tau_h]$. Existence of a maximum follows since $r(\cdot, \theta_h)$ is continuous. Consider $\theta_l \in [\tau_l, \tau_h]$. If $\theta_l \notin K$ and θ_l is a critical point, then $r_{\theta_l} = 0$ and so by Lemma 12, $r_{\theta_l \theta_l} < 0$. Thus θ_l is a strict local maximum.

Assume $\theta_l \in K$, and that θ_l is a critical point. Then, since $\pi_a(\theta_l, \gamma(\theta_l, \tilde{\kappa}), \tilde{v}) \geq 0$ and since a^{-n} jumps upwards at θ_l, we have that $r^-_{\theta_l}(\theta_l, \theta_h) \geq r^+_{\theta_l}(\theta_l, \theta_h)$. If $r^-_{\theta_l} < 0$ then $r^+_{\theta_l} < 0$, contradicting that θ_l is a critical point. Thus, $r^-_{\theta_l} \geq 0$. If $r^-_{\theta_l} > 0$, then, $r(\theta_l, \theta_h) > r(\theta'_l, \theta_h)$ for all θ'_l in a neighborhood to the left of θ_l. If instead $r^-_{\theta_l} = 0$, then, by (30) applied to the rectangle $\tilde{R}$ to the left of (θ_l, θ_h), $r(\cdot, \theta_h)$ is strictly concave on a neighborhood to the left of θ_l, and so, since $(r|_{\tilde{R}})_{\theta_l} = r^-_{\theta_l} = 0$, $r(\cdot, \theta_h)$ is strictly increasing on that neighborhood. Thus, again, $r(\theta_l, \theta_h) > r(\theta'_l, \theta_h)$ for all θ'_l in a neighborhood to the left of θ_l. Arguing similarly, $r(\theta_l, \theta_h) > r(\theta'_l, \theta_h)$ for all θ'_l in a neighborhood to the right of θ_l, and θ_l is again a strict local maximum. It follows that $r(\cdot, \theta_h)$ is strictly single-peaked on $[\tau_l, \tau_h]$, and hence has a single optimum on $[\tau_l, \tau_h]$. □.

Lemma 17 *The function ψ is continuous on $\left[\theta^x, \bar{\theta}\right]$. On D, λ is continuous as well.*

Proof Since L_N and L_S are strictly decreasing and continuous, the correspondence $\Theta(\cdot)$ is nonempty, compact-valued, and continuous, and so by the Theorem of the Maximum, ψ is continuous, and the set of maximizers of $r(\cdot, \theta_h)$ is upper hemicontinuous in θ_h. But then, since λ is single-valued on D by Lemma 16, it is continuous as a function on D. □

Our next lemma shows that the set of θ_h where profits are strictly positive is an interval.

Lemma 18 *The set $D = \{\theta_h > \theta^x | \psi(\theta_h) > 0\}$ is an interval.*

Proof Let E be the set on which $\theta > \theta^x$ and $v_*(\theta) > v^{-n}(\theta)$, where recall that $v_*(\theta) = B(\alpha_*(\theta)) - c(\alpha_*(\theta), \theta)$. By stacking and $C1$, v^{-n}, which has slope a^{-n}, is strictly steeper than v_θ, which has slope $\gamma(\theta, H(\theta))$, above θ^x. Hence E is an interval $(\theta^x, \bar{E})$. Our first step is to show that $E \subseteq D$. Our second step that for any $\theta_h > \bar{E}$, if $\theta_h \in D$, then an interval to the left of θ_h is also in D. Our third step is to show that since θ_h was an arbitrary point in D above $\bar{E}$, it follows that D is an interval as claimed.

Step 1 Choose $\theta_h \in E$. We wish to show $\theta_h \in D$. To do so, set α constant at $\alpha_*(\theta_h)$ and set $v(\theta_h) = v^{-n}(\theta_h)$. Then, by stacking and $C1$, (α, v) hires some non-empty interval of types $(\hat{\theta}_l, \theta_h)$, and earns $v_*(\theta_h) - v^{-n}(\theta_h) > 0$ on each type hired, since profits are constant by (10) using that the action is constant. *A fortiori*, $r(\hat{\theta}_l, \theta_h) > 0$. If we knew $\hat{\theta}_l \in \Theta(\theta_h) = [\tau_l, \tau_h]$, we would thus have shown $\theta_h \in D$, and hence $E \subseteq D$.

Let us first see that $\hat{\theta}_l \geq \tau_l$. Assume by way of contradiction that $\tau_l > \hat{\theta}_l$. Note that for $x < \theta_h$,

$$\frac{\partial}{\partial x}\left(v^{-n}(\theta_h) - v^{-n}(x) - \int_x^{\theta_h} \gamma(\theta_h, H(\theta_h))d\tau\right) = -a^{-n}(x) + \gamma(\theta_h, H(\theta_h)) > 0,$$

where the inequality uses $C1$, stacking, and that a^{-n} is increasing. Thus, by definition of $\hat{\theta}_l$,

$$\begin{aligned} 0 &= v^{-n}(\theta_h) - v^{-n}(\hat{\theta}_l) - \int_{\hat{\theta}_l}^{\theta_h} \gamma(\theta_h, H(\theta_h))d\tau \\ &< v^{-n}(\theta_h) - v^{-n}(\tau_l) - \int_{\tau_l}^{\theta_h} \gamma(\theta_h, H(\theta_h))d\tau \\ &< v^{-n}(\theta_h) - v^{-n}(\tau_l) - \int_{\tau_l}^{\theta_h} \gamma(\tau, H(\theta_h))d\tau \\ &= z\left(\tau_l, \theta_h, H\left(\theta_h\right)\right) = 0, \end{aligned}$$

where the second inequality uses that $\gamma\left(\cdot, H\left(\theta_h\right)\right)$ is strictly increasing, and the last two equalities use the definition of $\Theta\left(\theta_h\right)$, where we use that since $\tau_l > \hat{\theta}_l \geq \underline{\theta}$, the pair (τ_l, θ_h) is on L_S. This is a contradiction, and hence we have $\hat{\theta}_l \geq \tau_l$.

Assume that $\hat{\theta}_l > \tau_h$. We claim that in this case, $r\left(\cdot, \theta_h\right)$ is decreasing on $[\tau_h, \hat{\theta}_l]$. To see this, recall that $r\left(\cdot, \theta_h\right)$ is continuous and differentiable almost everywhere. Let $\tilde{\theta} \leq \hat{\theta}_l$ be given by

$$\tilde{\theta} = \inf\left\{\theta | r \text{ is decreasing on } (\theta, \hat{\theta}_l)\right\}, \tag{31}$$

and assume that $\tilde{\theta} > \tau_h$, noting that $\tilde{\theta} \leq \hat{\theta}_l$. But then, since r is decreasing on $(\theta, \hat{\theta}_l)$, and since $r(\hat{\theta}_l, \theta_h) > 0$, it follows that $r(\tilde{\theta}, \theta_h) > 0$, and so, since r is continuous, $r(\cdot, \theta_h) > 0$ on an interval to the left of $\tilde{\theta}$. But then, by Lemma 15, $r_{\theta_l}(\cdot, \theta_h) < 0$ almost everywhere on this interval as well, contradicting the definition of $\tilde{\theta}$. Thus $\tilde{\theta} \leq \tau_h$. But then, $\psi\left(\theta_h\right) \geq r\left(\tau_h, \theta_h\right) > 0$, and so once again $\theta_h \in D$, and we have established that $E \subseteq D$.

Step 2 Let $\theta_h > \bar{E}$ with $\theta_h \in D$. We wish to show that for θ_h' in an interval to the left of θ_h, θ_h' is also in D, with profits $\psi\left(\theta_h'\right) \geq \psi\left(\theta_h\right)$. To do so, note first that since $\theta_h > \bar{E}$, $\pi(\theta_h, \tilde{s}\left(\lambda\left(\theta_h\right), \theta_h\right)) < 0$, and so, since $\psi(\theta_h) > 0$, there is $\theta < \theta_h$ such that

$$0 > (\pi(\theta, \tilde{s}\left(\lambda\left(\theta_h\right), \theta_h\right))_\theta \underset{s}{=} \pi_a(\theta, \tilde{s}\left(\lambda\left(\theta_h\right), \theta_h\right))$$

where the equality of sign uses (10). Thus $\tilde{\kappa} < H(\theta_h)$ and so $(\lambda(\theta_h), \theta_h)$ is strictly above L_S. But then, for small $\varepsilon > 0$, $\lambda(\theta_h) \in \Theta(\theta_h - \varepsilon)$, and thus

$$\begin{aligned} \psi_{\theta_h}^{-}(\theta_h) &= \lim_{\varepsilon \downarrow 0} \frac{r(\lambda(\theta_h), \theta_h) - r(\lambda(\theta_h - \varepsilon), \theta_h - \varepsilon)}{\varepsilon} \\ &\leq \lim_{\varepsilon \downarrow 0} \frac{r(\lambda(\theta_h), \theta_h) - r(\lambda(\theta_h), \theta_h - \varepsilon)}{\varepsilon} \\ &= r_{\theta_h}^{-}(\lambda(\theta_h), \theta_h). \end{aligned}$$

But, since $\tilde{\kappa} < H(\theta_h)$, $\pi_a(\theta_h, \tilde{s}(\lambda(\theta_h), \theta_h)) < 0$, and thus since $\pi(\theta_h, \tilde{s}(\lambda(\theta_h), \theta_h)) < 0$ as well, by (23), $r_{\theta_h}^{-} < 0$ and hence $\psi_{\theta_h}^{-}(\theta_h) < 0$. But then, an interval to the left of θ_h is also in D with $\psi \geq \psi(\theta_h)$ as required.

Step 3 Finally, let us show that D is an interval. This follows by a construction similar to that around (31), using Step 2. In particular choose any $\theta_h \in D$ with $\theta_h > \bar{E}$, and let

$$\hat{\theta} = \inf \left\{ \theta \geq \bar{E} | [\theta, \theta_h] \subseteq D, \, \psi(\theta) \geq \psi(\theta_h) \right\},$$

and assume $\hat{\theta} > \bar{E}$. Then, by continuity of ψ, $\psi(\hat{\theta}) \geq \psi(\theta_h) > 0$, and so by Step 2 some interval to the left of $\hat{\theta}$ is also in D, with profits at least $\psi(\theta_h)$, contradicting the definition of $\hat{\theta}$. Hence $\hat{\theta} = \bar{E}$, and all of $(\theta^x, \theta_h) \in D$. Since θ_h was an arbitrary element of D, D is an interval as claimed, and we are done. □

Since $r_{\theta_l} < 0$ on L_N, $(\lambda(\theta_h), \theta_h)$ is never on L_N. But, because we maximize first with respect to θ_l, it may well be that $(\lambda(\theta_h), \theta_h) \in L_S$. Our next result shows that if $(\lambda(\theta), \theta)$ is on L_S, then ψ is strictly increasing. Hence, no such point is a local maximum or minimum of ψ. As mentioned in the text, the basic idea is that by Lemma 15, $r_{\theta_h}(\theta_l, \theta_h) > 0$ anywhere near L_S. But, since L_S is decreasing, as one moves a little above L_S, the constraint on θ_l is relaxed. Hence, $\psi_{\theta_h} \geq r_{\theta_h}(\theta_l, \theta_h) > 0$. The proof accounts for the presence of kinks and hence points where ψ is non-differentiable.

Lemma 19 *Let $(\lambda(\theta_h), \theta_h) \in L_S$ with $\theta_h \in D$. Then, $\psi_{\theta_h}^{+}(\lambda(\theta_h), \theta_h) > 0$, and $\psi_{\theta_h}^{-}(\lambda(\theta_h), \theta_h) > 0$, and hence ψ is not critical at θ_h.*

Proof Let $(\theta_l, \theta_h) = (\lambda(\theta_h), \theta_h) \in L_S$ with $\theta_h \in D$. Since the set of kinks K is finite, there is $\delta > 0$ such that $(\theta_h - \delta, \theta_h) \cap K = \emptyset$, such that $(\theta_l, \theta_l + \delta) \cap K = \emptyset$, and, using continuity of $\tilde{\kappa}$, such that $\tilde{\kappa}(\theta_l + \delta, \theta_h) > H(\theta_l)$, so that all of $(\theta_l, \theta_l + \delta) \times (\theta_h - \delta, \theta_h)$ lies strictly below L_N. From Lemma 10, r_{θ_h} is continuous on $(\theta_l, \theta_l + \delta) \times (\theta_h - \delta, \theta_h)$, and by examination of (23) and by Lemma 15, it follows that r_{θ_h} is continuous on $X \equiv \{(\theta_l, \theta_l + \delta) \times (\theta_h - \delta, \theta_h)\} \cup \{(\theta_l, \theta_h)\}$. Further, since $\theta_h \in D$, $r(\theta_l, \theta_h) > 0$, and so by Lemma 15, $r_{\theta_h}(\theta_l, \theta_h) > 0$.

Note next that for each θ_h' such that $(\lambda(\theta_h'), \theta_h') \in X \cap \Theta$,

$$\begin{aligned} \psi^+_{\theta_h}(\theta_h') &= \lim_{\varepsilon \downarrow 0} \frac{\psi(\theta_h' + \varepsilon) - \psi(\theta_h')}{\varepsilon} \\ &\geq \lim_{\varepsilon \downarrow 0} \frac{r(\lambda(\theta_h'), \theta_h' + \varepsilon) - r(\lambda(\theta_h'), \theta_h')}{\varepsilon} \\ &= r_{\theta_h}(\lambda(\theta_h'), \theta_h'), \end{aligned} \tag{32}$$

where the inequality follows since for small ε, $\lambda(\theta_h')$ is feasible at $\theta_h' + \varepsilon$.[24] Thus, in particular, since since $r_{\theta_h}(\lambda(\theta_h), \theta_h) > 0$ by Lemma 15, $\psi^+_{\theta_h}(\theta_h) > 0$.

Finally, consider $\psi^-_{\theta_h}(\theta_h)$. Fix $0 < \rho < r_{\theta_h}(\lambda(\theta_h), \theta_h)$. Since r_{θ_h} is continuous on X, and using (32), there is $\hat{\varepsilon} > 0$ such that for all $\varepsilon \in [0, \hat{\varepsilon})$, $\psi^+_{\theta_h}(\theta_h - \varepsilon) > \rho$. Let us show that $\psi^-_{\theta_h}(\theta_h) \geq \rho$, for which it is sufficient that for each $\varepsilon \in [0, \hat{\varepsilon})$, $\psi(\theta_h) - \psi(\theta_h - \varepsilon) \geq \rho\varepsilon$. Fix $\varepsilon \in [0, \hat{\varepsilon})$. Then, for any $\tau < \theta_h$ where $\psi(\tau) - \psi(\theta_h - \varepsilon) \geq \rho(\tau - (\theta_h - \varepsilon))$, the same inequality holds on an interval to the right of τ by the definition of a right derivative and by definition of ρ. But then, since $\psi(\theta_h - \varepsilon) - \psi(\theta_h - \varepsilon) \geq \rho \cdot 0$, it follows that $\psi(\theta_h) - \psi(\theta_h - \varepsilon) \geq \rho\varepsilon$ by a construction similar to that around (31), and we are done. □

Given this result, we turn attention to places where the path described by λ does not lie on L_S. Let $D' = \{\theta_h \in D | (\lambda(\theta_h), \theta_h) \notin L_S\}$. Our next lemma shows that ψ_{θ_h} and r_{θ_h} agree on D'. As described in the text, this holds by what is essentially the Envelope Theorem.

Lemma 20 *For all $\theta_h \in D'$, $\psi^+_{\theta_h}(\theta_h) = r^+_{\theta_h}(\lambda(\theta_h), \theta_h)$ and $\psi^-_{\theta_h}(\theta_h) = r^-_{\theta_h}(\lambda(\theta_h), \theta_h)$.*

Proof Let $K_1 = (K \cap \lambda(D')) \cup \{\underline{\theta}\}$ and $K_2 = K \cap D'$. Thus, $r(\lambda(\cdot), \cdot)$ may be non-differentiable because either $\theta_h \in K_2$ or $\lambda(\theta_h) \in K_1$. There are thus several cases to consider.

Case 1 Consider first $\theta_h \in D'$ such that $\lambda(\theta_h) \notin K_1$ and $\theta_h \notin K_2$. Then, we are not on L_S by definition of D', and we are not on L_N since by Lemma 15, $r_{\theta_l}(\theta_l, \theta_h) < 0$ on L_N. Thus, since $\psi(\theta_h) = r(\lambda(\theta_h), \theta_h)$,

$$\psi_{\theta_h}(\theta_h) = r_{\theta_h}(\lambda(\theta_h), \theta_h), \tag{33}$$

by the Envelope Theorem.

Case 2 For any given $\theta_l \in K_1$, let $\underline{J}(\theta_l) = \min\{\theta_h | \lambda(\theta_h) = \theta_l\}$, and let $\bar{J}(\theta_l) = \max\{\theta_h | \lambda(\theta_h) = \theta_l\}$.[25] Let $J(\theta_l) = (\underline{J}(\theta_l), \bar{J}(\theta_l))$. Since λ is constant on $J(\theta_l)$, if $J(\theta_l)$ is non-empty, then for all $\theta_h \in J(\theta_l) \backslash K_2$, we have again have (33).

Case 3 Consider next $\theta_h \in (\{\underline{J}(\theta_l)\}_{\theta_l \in K_1} \cup \{\bar{J}(\theta_l)\}_{\theta_l \in K_1}) \backslash K$. Assume that $\theta_h = \underline{J}(\theta_l)$ for some $\theta_l \in K_1$ (the case where $\theta_h = \bar{J}(\theta_l)$ is similar). Then, $\psi_{\theta_h}(\theta_h') = r_{\theta_h}(\lambda(\theta_h'), \theta_h')$ for θ_h' on a

[24] That is, $\lambda(\theta_h') \in \Theta(\theta_h' + \varepsilon)$, since L_S is decreasing and $(\theta_l, \theta_l + \delta) \times (\theta_h - \delta, \theta_h)$ lies strictly below L_N.

[25] These correspond to the bottoms and tops of the vertical segments of the path in Figure 3.

neighborhood above θ_h by Case 2. For a neighborhood below θ_h, $\theta_h' \notin K$, since K is finite, and $\lambda(\theta_h') \notin K$ by definition of $\underline{J}(\theta_l)$ and again since K is finite. Hence $\psi_{\theta_h}(\theta_h') = r_{\theta_h}(\lambda(\theta_h'), \theta_h')$ by (33). But then, by continuity of r_{θ_h} on these neighborhoods, and by continuity of λ, $\psi_{\theta_h}(\theta_h) = r_{\theta_h}(\lambda(\theta_h), \theta_h)$ as well.

Case 4 Finally, consider $\theta_h \in K_2$. Since K is finite, on some neighborhood above θ_h, $\psi_{\theta_h} = r_{\theta_h}$ by the previous cases, and λ is continuous, and so

$$\psi^+_{\theta_h}(\theta_h) = \lim_{\varepsilon \downarrow 0} \frac{\psi(\theta_h + \varepsilon) - \psi(\theta_h)}{\varepsilon} = \lim_{\varepsilon \downarrow 0} \psi_{\theta_h}(\theta_h + \varepsilon) = \lim_{\varepsilon \downarrow 0} r_{\theta_h}(\lambda(\theta_h + \varepsilon), \theta_h + \varepsilon) = r^+_{\theta_h}(\lambda(\theta_h), \theta_h),$$

using L'Hôpital's rule at the second inequality, and where to justify the last equality, we note from (23) that r_{θ_h} does not depend on $a^{-n}(\theta_l)$, and so it does not matter whether or not $\lambda(\theta_h) \in K_1$. Similarly, $\psi^-_{\theta_h}(\theta_h) = r^-_{\theta_h}(\lambda(\theta_h), \theta_h)$. □

We are now ready to show that ψ has a unique maximum on D.

Lemma 21 *The function ψ is strictly single-peaked, and thus has a unique maximum, on D.*

Proof We will show first that if θ_h is a critical point, then it is a strict local maximum of ψ. By Lemma 19, any critical point of ψ is in D'. This will follow because Lemma 20, lets us relate the local concavity properties of ψ to those we establish for r in Lemma 12. We go through the same four cases as in Lemma 20.

Case 1 Consider first $\theta_h \in D'$ such that $\lambda(\theta_h) \notin K_1$ and $\theta_h \notin K_2$. Then, since $\underline{\theta} \in K_1$, $(\lambda(\theta_h), \theta_h) \in \Theta$, and so Lemma 12 applies, and thus, by the Implicit Function Theorem, $\lambda_{\theta_h} = -r_{\theta_l\theta_h}/r_{\theta_l\theta_l}$, where by Lemma 12 $r_{\theta_l\theta_l} < 0$. Since (33) holds on a neighborhood of θ_h,

$$\begin{aligned} \psi_{\theta_h\theta_h}(\theta_h) &= r_{\theta_h\theta_l}(\lambda(\theta_h), \theta_h)\lambda_{\theta_h}(\theta_h) + r_{\theta_h\theta_h}(\lambda(\theta_h), \theta_h) \\ &= -\frac{(r_{\theta_l\theta_h}(\lambda(\theta_h), \theta_h))^2}{r_{\theta_l\theta_l}(\lambda(\theta_h), \theta_h)} + r_{\theta_h\theta_h}(\lambda(\theta_h), \theta_h) \\ &= \frac{1}{r_{\theta_l\theta_l}(\lambda(\theta_h), \theta_h)} \left(r_{\theta_l\theta_l}(\lambda(\theta_h), \theta_h) r_{\theta_h\theta_h}(\lambda(\theta_h), \theta_h) - (r_{\theta_l\theta_h}(\lambda(\theta_h), \theta_h))^2 \right). \end{aligned} \tag{34}$$

But then, if θ_h is a critical point, so that $\psi_{\theta_h}(\theta_h) = 0$, then $r_{\theta_l} = r_{\theta_h} = 0$ at $(\lambda(\theta_h), \theta_h)$, and so by Lemma 12 $r_{\theta_l\theta_l}r_{\theta_h\theta_h} - r^2_{\theta_l\theta_h} > 0$. Hence, $\psi_{\theta_h\theta_h}(\theta_h) < 0$, and θ_h is a strict local maximum of ψ.

Case 2 Consider $\theta_h \in D'$ where $\theta_h \notin K_2$ but for some $\theta_l \in K_1$, $\lambda(\theta_h) \in J(\theta_l)$. Then, since $J(\theta_l) \backslash K_2$ is open, by Case 2 of Lemma 20, (33) holds on a neighborhood of θ_h, and so, since λ is constant on $J(\theta_l)$, $\psi_{\theta_h\theta_h}(\theta_h) = r_{\theta_h\theta_h}(\theta_l, \theta_h)$. If $\theta_h \le \theta_T$, so that $(\lambda(\theta_h), \theta_h) \in \Theta$, then by Lemma 12, if $\psi_{\theta_h}(\theta_h) = 0$, then $\psi_{\theta_h\theta_h}(\theta_h) < 0$, so θ_h is a strict local maximum of ψ. Assume that $\theta_h \ge \theta_T$, so that $\lambda(\theta_h) = \underline{\theta}$ and $\tilde{\kappa}(\lambda(\theta_h), \theta_h) = 0$. Trace the derivation of $r_{\theta_h\theta_h}$ in the proof of Lemma 12 up through (29) with $\tilde{\kappa}$ replaced by 0, and note that this part of the proof relies on $\tilde{v}(\theta_h) = v^{-n}(\theta_h)$

but not on $\tilde{v}(\theta_l) = v^{-n}(\theta_l)$. It follows that where $r_{\theta_h}(\underline{\theta}, \theta_h) = 0$, $\psi_{\theta_h\theta_h}(\underline{\theta}, \theta_h) = r_{\theta_h\theta_h}(\underline{\theta}, \theta_h) < 0$, and again, θ_h is a strict local maximum of ψ.

Case 3 Consider next $\theta_h \in (\{\underline{J}(\theta_l)\}_{\theta_l \in K_1} \cup \{\bar{J}(\theta_l)\}_{\theta_l \in K_1}) \backslash K$. Assume that $\theta_h = \underline{J}(\theta_l)$ for some $\theta_l \in K_1$ (the other case is similar), and assume that $\psi_{\theta_h}(\theta_h) = 0$. Then by Case 2, ψ is strictly concave on a neighborhood just above θ_h, while by Case 1, ψ is strictly concave on a neighborhood just below θ_h. Hence, again, θ_h is a strict local maximum of ψ.

Case 4 Finally, consider $\theta_h \in K_2 = K \cap D'$. Since $\tilde{\kappa} \in [0, 1]$, and since $\theta_h > \theta^x$, we have that $a^{-n} - \gamma$ is positive and bounded away from 0 and ∞ on a neighborhood of θ_h by stacking and $C1$. At any point θ'_h of continuity of a^{-n}, and repeating (23) for convenience,

$$\frac{\psi_{\theta_h}(\theta'_h)}{h(\theta'_h)} = \frac{r_{\theta_h}(\lambda(\theta'_h), \theta'_h)}{h(\theta'_h)} = \pi(\theta'_h, \gamma(\cdot, \tilde{\kappa}), v^{-n}(\theta'_h)) + \pi_a(\theta'_h, \gamma(\cdot, \tilde{\kappa}), v^{-n}(\theta'_h)) \left(a^{-n}(\theta'_h) - \gamma(\theta'_h, \tilde{\kappa})\right), \tag{35}$$

where we recall that $\tilde{\kappa}$ is continuous, and hence so is $\gamma^n(\cdot, \tilde{\kappa})$, and that v^{-n} is also continuous, and hence so are π and π_a. Thus any discontinuity in ψ_{θ_h} at θ_h is driven by an upward jump of a^{-n} at θ_h and, since $\pi_a(\theta_h, \gamma(\cdot, \tilde{\kappa}), v^{-n}(\theta_h)) \leq 0$ (since $\tilde{\kappa} \leq H(\theta_h)$), for there to be a discontinuity, we must have $\pi_a < 0$.

If $\pi \leq 0$, then, by (35) both $\psi^+_{\theta_h}(\theta_h)$ and $\psi^-_{\theta_h}(\theta_h)$ are strictly negative, and θ_h is not a critical point. If $\pi > 0$,then since a^{-n} jumps up at θ_h, we have $\psi^-_{\theta_h} > \psi^+_{\theta_h}$. Assume that θ_h is a critical point, so that $\psi^-_{\theta_h}\psi^+_{\theta_h} \leq 0$. If $\psi^-_{\theta_h} > 0 > \psi^+_{\theta_h}$, then θ_h is a strict local maximum of ψ. If $\psi^+_{\theta_h} = 0$, then, first, $\psi^-_{\theta_h} > 0$, and, second, from the previous cases, $\psi_{\theta_h\theta_h} < 0$ for all θ on a neighborhood to the right of θ_h. Similarly if $\psi^-_{\theta_h} = 0$, then $\psi^+_{\theta_h} < 0$, and $\psi_{\theta_h\theta_h} < 0$ for all θ on a neighborhood to the left of θ_h. In each case θ_h is again a strict local maximum of ψ.

Thus, if $\theta_h \in D$ is a critical point of ψ, then θ_h is a strict local maximum of ψ. Since D is a non-empty interval, ψ is strictly single-peaked on D, and so has a unique maximum, and any critical point of ψ in D is that maximum. □

Finally, let us connect the maximization of ψ to that of r.

Lemma 22 *Let θ^*_h be the unique maximizer of ψ. Then, the unique maximizer of r is $(\lambda(\theta^*_h), \theta^*_h)$.*

Proof Let $(\theta^{**}_l, \theta^{**}_h) \in \arg\max_{\{(\theta_l,\theta_h)|\theta_h \geq \theta_l\}} r(\theta_l, \theta_h)$. Since D is non-empty, $r(\theta^{**}_l, \theta^{**}_h) > 0$, and hence $\theta^{**}_l < \theta^{**}_h$, and $\theta^{**}_h \in D$. By Lemma 15, $(\theta^{**}_l, \theta^{**}_h) \in \Theta \cup A$, and so $\theta^{**}_l \in \Theta(\theta^{**}_h)$. Hence by Lemma 16, $\theta^{**}_l = \lambda(\theta^{**}_h)$. Since $(\theta^{**}_l, \theta^{**}_h)$ is optimal and since the constraint $\theta_h \geq \theta_l$ is slack, we must have $r^+_{\theta_h}(\lambda(\theta^{**}_h), \theta^{**}_h) \leq 0$ and $r^-_{\theta_h}(\lambda(\theta^{**}_h), \theta^{**}_h) \geq 0$. But then, by Lemma 20, $\psi^+_{\theta_h}(\theta^{**}_h) \leq 0$ and $\psi^-_{\theta_h}(\theta^{**}_h) \geq 0$, and so by Lemma 21 $\theta^{**}_h = \theta^*_h$, and we are done. □

9.4 Proofs for Section 6.6

Proof of Theorem 4 Let $\hat{s}$ satisfy stacking, PS, IO and OB. Fix n and let $\hat{s}^n = (\hat{\alpha}, \hat{v})$, with associated $\hat{\kappa}$. By IO, $(\hat{\alpha}, \hat{v})$ satisfies $C1$ on (θ_l, θ_h). But then, by IO, if $n < N$, then $\pi_a(\theta_h, \hat{\alpha}, \hat{v}) < 0$, and by $C1$ and stacking, $a^{-n}(\theta_h) - \hat{\alpha}(\theta_h) > 0$. Hence, by (9), $\pi(\theta_h, \hat{\alpha}, \hat{v}) > 0$. Similarly, $\pi(\theta_l, \hat{\alpha}, \hat{v}) > 0$ if $n > 1$. But then, since by Lemma 2 profits are strictly single-peaked with maximum at θ_0 solving $H(\theta_0) = \hat{\kappa}$, $\pi(\theta, \hat{\alpha}, \hat{v}) > 0$ for all $\theta \in [\theta_l, \theta_h]$. Thus $\hat{v}(\theta) < v_*(\theta)$, so that $(\hat{\alpha}, \hat{v})$ satisfies $C2$ on $[\theta_l, \theta_h]$.

Let us first re-define $(\hat{\alpha}, \hat{v})$ outside of $[\theta_l, \theta_h]$ to satisfy $C1$ and $C2$ there as well. Set

$$\alpha(\theta) = \begin{cases} \min\{\gamma(\theta, 0), \hat{\alpha}(\theta_l)\} & \theta < \theta_l \\ \hat{\alpha}(\theta) & \theta \in [\theta_l, \theta_h] \\ \max\{\gamma(\theta, 1), \hat{\alpha}(\theta_h)\} & \theta > \theta_h \end{cases},$$

and set $v(\theta) = \hat{v}(\theta_l) + \int_{\theta_l}^{\theta} \alpha(\tau) d\tau$ for all θ. That is, modify $(\hat{\alpha}, \hat{v})$ such that actions and surplus are unchanged in $[\theta_l, \theta_h]$, and modified outside of $[\theta_l, \theta_h]$ to ensure that $C1$ holds while respecting monotonicity. Note that $\hat{\alpha}(\theta_h) = \gamma(\theta_h, \hat{\kappa}) \geq \gamma(\theta_h, 1)$, and so no discontinuity is introduced at θ_h, and similarly at θ_l. By stacking, it remains the case that (α, v) is single-dominant on $[\theta_l, \theta_h]$, and so, since (α, v) and $(\hat{\alpha}, \hat{v})$ agree on $[\theta_l, \theta_h]$, (α, v) and $(\hat{\alpha}, \hat{v})$ are essentially equivalent.

To show that $C2$ is now satisfied for $\theta \notin [\theta_l, \theta_h]$, assume $(\theta_h, \bar{\theta}]$ is non-empty (the argument when $[\underline{\theta}, \theta_l)$ is non-empty is the same). On the interval where $\alpha(\cdot) = \hat{\alpha}(\theta_h)$, $(\pi(\theta, \alpha, v))_\theta = \pi_a(\theta, \alpha, v)\alpha_\theta(\theta) = 0$ by (10). Where $\alpha(\cdot) = \gamma(\cdot, 1)$, $(\pi(\theta, \alpha, v))_\theta = \pi_a(\theta, \gamma(\cdot, 1), v)\gamma_\theta(\theta, 1) \geq 0$, using that $\gamma_\theta(\theta, 1) > 0$, that $\gamma(\theta, 1) \leq \gamma(\theta, H(\theta)) = \alpha_*(\theta)$, and that π is strictly concave in a, and so $\pi_a(\theta, \gamma(\cdot, 1), v) \geq 0$. Thus, $\pi(\theta, \alpha, v) \geq \pi(\theta_h, \alpha, v) > 0$ for all $\theta > \theta_h$, and so $v(\theta) < v_*(\theta)$ and $C2$ is satisfied.

Construct the strategy profile s by performing the above process for each n. Then OB continues to hold for all n, since for each of n's opponents, $\hat{\alpha}$ and α agree on $[\theta_l, \theta_h]$, and since both the modified and original action profiles of n's opponents are continuous. Let us show that s is a Nash equilibrium. Fix $n \notin \{1, N\}$. Assume first that Assumption 1 holds. By the argument in the first paragraph of this proof, $\theta_h \in D$. Since by PS, $\underline{\theta} < \theta_l < \theta_h < \bar{\theta}$, it follows that $z(\theta_l, \theta_h, \tilde{\kappa}(\theta_l, \theta_h)) = 0$, where $\tilde{\kappa}(\theta_l, \theta_h) \in (H(\theta_l), H(\theta_h))$ by IO, and so we have $\theta_l \in \Theta(\theta_h)$. But then, since $r_{\theta_l}(\theta_l, \theta_h) = 0$ by OB, we must have $\theta_l = \lambda(\theta_h)$ by Lemma 16. But then, again by OB, $0 = r_{\theta_h}(\theta_l, \theta_h) = r_{\theta_h}(\lambda(\theta_h), \theta_h) = \psi_{\theta_h}(\theta_h)$, where the third equality is by Lemma 20. Finally, since by Lemma 21 ψ is strictly single-peaked on the interval D, we have $\theta_h = \theta_h^*$ by Lemma 22. Thus, s^n is a best response to s^{-n} by Corollary 2.

If Assumption 1 fails, then recall from the end of Section 6.5 that $\tilde{\lambda}$ is the analogue to λ. So, we argue first that $\theta_h \in \tilde{\Theta}(\theta_l)$, then by the analogue to Lemma 16 that $\theta_h = \tilde{\lambda}(\theta_l)$, and then by the analogue to Lemma 20 that by OB $0 = r_{\theta_l}(\theta_l, \theta_h) = r_{\theta_l}(\theta_l, \tilde{\lambda}(\theta_l)) = \tilde{\psi}_{\theta_l}(\theta_l)$. But then, since $\tilde{\psi}$

is strictly single-peaked on $\tilde{D}$, we have $\theta_l = \tilde{\theta}_l^*$, and again s^n is a best response to s^{-n}.

Consider $n = 1$. Then, $\kappa^1 = 0$ by IO, and $\theta_l = \underline{\theta}$ by PS. But, since $\kappa^1 = 0$, and since, as argued in the first paragraph of this proof, $\pi(\theta_h, \hat{\alpha}, \hat{v}) > 0$, it follows by Lemma 2 that $\pi(\theta, \hat{\alpha}, \hat{v}) > 0$ for all $\theta < \theta_h$. Hence, since $\pi_a(\theta, \hat{\alpha}, \hat{v}) < 0$ for all $\theta < \theta_h$, we have $r_{\theta_l} < 0$, and so $\underline{\theta} = \lambda(\theta_h)$. But then, by OB, $0 = r_{\theta_h}(\underline{\theta}, \theta_h) = r_{\theta_h}(\lambda(\theta_h), \theta_h) = \psi_{\theta_h}(\theta_h)$, and again $\theta_h = \theta_h^*$, and s^1 is a best response to s^{-1}.

Finally, consider $n = N$. Then, since $\kappa^N = 1$ by IO, it follows that $\theta_h = \bar{\theta} = \tilde{\lambda}(\theta_l)$. Thus, by OB, $0 = r_{\theta_l}(\theta_l, \bar{\theta}) = r_{\theta_l}(\theta_l, \tilde{\lambda}(\theta_l)) = \tilde{\psi}_{\theta_l}(\theta_l)$, and so $\theta_l = \tilde{\theta}_l^*$, and s^N is a best response to s^{-N}. □

9.5 Proofs for Section 6.7

We begin by defining three further restrictions on strategies that will turn out not to bind in equilibrium, but that help us towards compactness and continuity.

Recall that $BR(s^{-n}) = \arg\max_{s^n \in S^n} \Pi^n(s^n, s^{-n})$. Let

$$\eta = \max\{\gamma^N(\overline{\theta}, 0), \max_{n,\theta,\kappa\in[0,1]} \gamma^n_\theta(\theta, \kappa)\}. \tag{36}$$

Since $\pi^n_{aa} = B^n_{aa}$, and by definition of γ,

$$\gamma^n_\theta(\theta, \kappa) = \frac{1}{B^n_{aa}(\gamma^n(\theta, \kappa))} \left(\left(\frac{\kappa - H(\theta)}{h(\theta)} \right)_\theta - 1 \right).$$

Since $B^n_{aa} < 0$ is continuous, it is bounded away from zero on the set of actions $\{\gamma^n(\theta, \kappa) \,|\theta \in [\underline{\theta}, \overline{\theta}], \kappa \in [0, 1]\}$, which is compact since γ^n is continuous. But then, since h is $\mathcal{C}^1$ and bounded away from 0, η is well-defined and finite.

Since $\gamma^n(\theta, \kappa) \leq \gamma^N(\overline{\theta}, 0)$ for all n, θ, and $\kappa \in [0, 1]$, η is a bound on the highest value and slope of any γ satisfying $C1$. We will thus bound the slopes of our allowable action profiles by η.

C3 $0 \leq \alpha^n(\theta') - \alpha^n(\theta) \leq \eta(\theta' - \theta)$ for all θ, θ' with $\theta' > \theta$.

Next, let

$$\beta = \min_{n,\theta_h,\kappa\in[0,1]} \left(\pi^n_a(\theta_h, \gamma^n(\cdot, \kappa), 0)\gamma^N(\theta_h, 0) + \pi^n(\theta_h, \gamma^n(\cdot, \kappa), 0) \right)$$

where $\beta > -\infty$ since each relevant object is continuous and hence bounded on the compact choice set. We will see that anywhere that (9) holds, $v^n(\bar{\theta}) \geq \beta$, motivating our next restriction.

C4 $v^n(\bar{\theta}) \geq \beta$.

For each n, let S^n_R be the subset of S^n such that $C1$–$C4$ hold. Let $S_R = \times_{n'} S^{n'}_R$, and $S^{-n}_R = \times_{n' \neq n} S^{n'}_R$.

Lemma 23 *Fix $s^{-n} \in S_R^{-n}$. Then $BR^n(s^{-n}) \cap S_R^n$ is nonempty.*

Proof Fix n, and fix $\hat{s}^n \in BR^n(s^{-n})$, where we note that $BR^n(s^{-n})$ is non-empty since r has a maximizer and using Corollary 2. Further, by that Corollary, and using stacking, $\hat{s}^n$ is single-dominant on some region $[\theta_l, \theta_h]$, and has the form $(\hat{\alpha}, \hat{v})$, where $\hat{\alpha} = \gamma(\cdot, \kappa)$ on $[\theta_l, \theta_h]$, where $\kappa \in [H(\theta_l), H(\theta_h)]$, and where $C1$ and $C2$ are satisfied on $[\theta_l, \theta_h]$. Let (α, v) be defined from $(\hat{\alpha}, \hat{v})$ as in the proof of Theorem 4, so that as shown there, $C1$ and $C2$ are satisfied on $[\underline{\theta}, \bar{\theta}]$. By stacking, and using that for $n' \neq n$, $C1$ and $C2$ are satisfied by assumption, it remains the case that (α, v) is single-dominant on $[\theta_l, \theta_h]$, and since (α, v) and $(\hat{\alpha}, \hat{v})$ agree on $[\theta_l, \theta_h]$, it follows that $(\alpha, v) \in BR(s^{-n})$. Condition $C3$ holds by construction.

To show that $C4$ holds, assume by way of contradiction, that $v(\bar{\theta}) < \beta$. Then, since $v^{n'}(\bar{\theta}) \geq \beta$ for each $n' \neq n$, $\theta_h < \bar{\theta}$, and so by (23), if we let $\bar{a} = \lim_{\theta_h' \downarrow \theta_h} a^{-n}(\theta_h')$, then, by Corollary 2, since (θ_l, θ_h) maximized r,

$$0 \geq \frac{r^+_{\theta_h}(\theta_l, \theta_h)}{h(\theta_h)} = \pi_a(\theta_h, \gamma(\cdot, \kappa), v)(\bar{a} - \gamma(\theta_h, \kappa)) + \pi(\theta_h, \gamma(\cdot, \kappa), v).$$

But, since s^n is a best response, it follows from Proposition 3 and continuity of π, γ, and v that $\pi(\theta_h, \gamma(\cdot, \kappa), v) \geq 0$. By $C1$ and $C2$ for $n' \neq n$, and stacking, $\bar{a} - \gamma(\theta_h, \kappa) > 0$. Hence $\pi_a(\theta_h, \gamma(\cdot, \kappa), v) \leq 0$, and so we have

$$\begin{aligned} 0 &\geq \pi_a(\theta_h, \gamma(\cdot, \kappa), v)\gamma^N(\theta_h, 0) + \pi(\theta_h, \gamma(\cdot, \kappa), v) \\ &> \pi_a(\theta_h, \gamma(\cdot, \kappa), 0)\gamma^N(\theta_h, 0) + \pi(\theta_h, \gamma(\cdot, \kappa), 0) - \beta \\ &\geq 0, \end{aligned}$$

where the first inequality uses $\bar{a} - \gamma(\theta_h, \kappa) \leq \bar{a} \leq \gamma^N(\theta_h, 0)$, the second uses monotonicity of v and $v(\bar{\theta}) < \beta$, and the last inequality uses the definition of β. This is a contradiction, and hence $v(\bar{\theta}) \geq \beta$ as required. Since (α, v) is a best response and satisfies $C1 - C4$, we are done. □

Proof of Theorem 5 Let us now prove that the game $(S^n, \Pi^n)_{n=1}^N$ has a pure-strategy equilibrium. It is enough to show that $(S_R^n, \Pi^n)_{n=1}^N$ has a pure-strategy equilibrium: By Lemma 23 $BR^n(s^{-n}) \cap S_R^n$ is nonempty, and so in a Nash equilibrium of $(S_R^n, \Pi^n)_{n=1}^N$, each player is playing an element of $BR^n(s^{-n})$, and we have a Nash equilibrium of $(S^n, \Pi^n)_{n=1}^N$.

The set of continuous functions from $[\underline{\theta}, \bar{\theta}]$ to $\mathbb{R}$, endowed with the sup norm $\|\cdot\|_\infty$, is a Banach space, and thus S_R^n, with norm $\|(\alpha^n, v^n)\| = \|\alpha^n\|_\infty + \|v^n\|_\infty$ is a subset of a Banach space. Similarly S_R with norm $\sum_n \|(\alpha^n, v^n)\|$ is a subset of a Banach space.

Let us show that for each n, the set S_R^n is nonempty, convex, and compact. To see that S_R^n is nonempty, we will argue that $(\alpha_*^n, v_*^n) \in S_R^n$. Note that $C2$ is immediate, and that $C1$ follows because $\alpha_*^n(\theta) = \gamma^n(\theta, H(\theta))$. But then, by definition of η,

$$(\alpha_*^n(\theta))_\theta = \gamma_\theta^n(\theta, H(\theta)) + \gamma_\kappa^n(\theta, H(\theta))h(\theta) < \gamma_\theta^n(\theta, H(\theta)) \leq \eta,$$

using that $\gamma_\kappa^n < 0$, and so $C3$ follows. To see $C4$, note that since $\alpha_*^n(\bar{\theta}) = \gamma^n(\bar{\theta}, 1)$, it follows that $\pi_a^n(\bar{\theta}, \gamma^n(\bar{\theta}, 1), 0) = 0$, and hence

$$\pi_a^n(\bar{\theta}, \gamma^n(\bar{\theta}, 1), 0)\gamma^N(\theta_h, 0) + \pi^n(\bar{\theta}, \gamma^n(\bar{\theta}, 1), 0) = v_*^n(\bar{\theta}),$$

and thus $v_*^n(\bar{\theta}) \geq \beta$. Thus, S_R^n is nonempty.

To prove convexity of S_R^n, let (α_1^n, v_1^n) and $(\alpha_2^n, v_2^n) \in S_R^n$, let $\delta \in [0, 1]$, and let $(\alpha_3^n, v_3^n) = (\delta\alpha_1^n + (1-\delta)\alpha_2^n, \delta v_1^n + (1-\delta)v_2^n)$. Then, (α_3^n, v_3^n) satisfies the integral condition $v_3^n(\theta) = v_3^n(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \alpha_3^n(\tau)\, d\tau$, since integration is a linear operator, so that $(\alpha_3^n, v_3^n) \in S^n$, and it is direct that (α_3^n, v_3^n) satisfies $C1$–$C4$.

To prove compactness, let $(\alpha_k^n, v_k^n)_{k=1}^\infty$ be a sequence of elements of S_R^n. Then, by $C1$ and the definition of η, we have $\alpha_k^n(\underline{\theta}) \geq 0$ and $\alpha_k^n(\bar{\theta}) \leq \eta$. Hence, since $C3$ is satisfied by α_k^n for each k, it follows by the Arzela-Ascoli Theorem (e.g., Rudin (1987), Theorem 11.28, p. 245) that there exists α^n satisfying $C1$ and $C3$ and a subsequence along which $\|\alpha_k^n - \alpha^n\|_\infty \to 0$. Note that α^n is increasing and has range contained in $[0, \eta]$, and so is integrable. Since $v_k^n(\bar{\theta})$ lies in a compact set by $C2$ and $C4$, we can also, by taking a further subsequence and re-indexing, assume that along the chosen subsequence $v_k^n(\bar{\theta}) \to \bar{v}$, for some $\bar{v}$. For each $\theta \in [\underline{\theta}, \bar{\theta}]$, define $v^n(\theta) = \bar{v} - \int_\theta^{\bar{\theta}} \alpha^n(\tau) d\tau$. We claim that (i) along the same subsequence, $\|v_k^n - v^n\|_\infty \to 0$, and (ii) $(\alpha^n, v^n) \in S_R^n$. To see (i), note that for each θ and k, $v_k^n(\theta) = v_k^n(\bar{\theta}) - \int_\theta^{\bar{\theta}} \alpha_k^n(\tau)d\tau$, and hence

$$|v^n(\theta) - v_k^n(\theta)| \leq \left|v_k^n(\bar{\theta}) - \bar{v}\right| + \int_\theta^{\bar{\theta}} |\alpha_k^n(\tau) - \alpha^n(\tau)|\, d\tau \leq \left|v_k^n(\bar{\theta}) - \bar{v}\right| + \left(\bar{\theta} - \underline{\theta}\right) \|\alpha_k^n - \alpha^n\|_\infty$$

and thus, since the last expression is independent of θ, $\|v_k^n - v^n\|_\infty \to 0$. To see (ii), note that we have already checked $C1$ and $C3$, and that weak inequalities are preserved under limits, and so $C2$ and $C4$ hold as well. Thus, S_R^n is sequentially compact and is a metric space, it is compact.

Since N is finite and, for each n, S_R^n is nonempty, convex, and compact, so is the product $S_R = \times_{i=1}^N S_R^n$. Fix $s \in S_R$, let $s_k \to s$, and fix n. Then, by stacking and since $s \in S_R$, there exist θ_l and θ_h such that $\varphi(\theta, s) = 1$ on (θ_l, θ_h) and $\varphi(\theta, s) = 0$ for $\theta \notin [\theta_l, \theta_h]$. But then, since for each n', $\|v_k^{n'} - v^{n'}\| \to 0$, and again using stacking, for any given $\delta > 0$, and for s' close enough to s, $\varphi(\theta, s') = 1$ on $[\theta_l + \delta, \theta_h - \delta]$ and $\varphi(\theta, s') = 0$ for $\theta \notin (\theta_l - \delta, \theta_h + \delta)$. Since $\|\alpha_k^n - \alpha^n\| \to 0$ as well, and since π is bounded and continuous, it follows that $\Pi^n(s_k) \to \Pi^n(s)$, and thus that Π^n is continuous on S_R.[26]

[26] Recall that without stacking, and outside of S_R, it is easy to construct examples where payoffs are discontinuous.

Fix n. Since Π^n is continuous on S_R, and since S_R^n is non-empty, compact, and independent of v^{-n} (and hence trivially continuous as a correspondence) the Theorem of the Maximum implies that $BR_R^n(s^{-n}) = \arg\max_{s^n \in S_R^n} \Pi^n(s^n, s^{-n})$ is non-empty and compact valued for each s^{-n}, and is upper hemicontinuous in s^{-n}.

Finally, let us show that $BR_R^n(s^{-n})$ is convex. Let $\hat{s}^n \in BR_R^n(s^{-n})$, with single- dominance region $(\hat{\theta}_l, \hat{\theta}_h)$. Then, by Corollary 2, $(\hat{\theta}_l, \hat{\theta}_h)$ maximizes r, and on $(\hat{\theta}_l, \hat{\theta}_h)$, $\hat{s}^n = \tilde{s}(\hat{\theta}_l, \hat{\theta}_h)$, and by Lemma 22, $(\hat{\theta}_l, \hat{\theta}_h) = (\lambda(\theta_h^*), \theta_h^*)$. Thus, any two elements of $BR_R^n(s^{-n})$ win for sure on $(\lambda(\theta_h^*), \theta_h^*)$ and agree with $\tilde{s}(\lambda(\theta_h^*), \theta_h^*)$ on $(\lambda(\theta_h^*), \theta_h^*)$, and lose for sure for $\theta \notin [\lambda(\theta_h^*), \theta_h^*)]$. But then, their convex combination does the same, and so is also a best response.

We have shown that S_R is a non-empty, compact, convex subset of a Banach space, and that the correspondence defined by $BR_R(s) \equiv BR_R^1(s^{-1}) \times \cdots \times BR_R^N(s^{-N})$ from S_R to S_R has a closed graph and nonempty convex values. Thus, by the Kakutani-Fan-Glicksberg Theorem (Aliprantis and Border (2006), Corollary 17.55, p. 583) BR_R has a fixed-point on S_R, and we are done. □

References

Aliprantis, C., and K. Border (2006): *Infinite Dimensional Analysis.* Springer Verlag, Berlin.

Biglaiser, G., and C. Mezzetti (1993): "Principals Competing for an Agent in the Presence of Adverse Selection and Moral Hazard," *Journal of Economic Theory*, 61(2), 302–330.

Chamberland, M. (2015): *Single Digits: In Praise of Small Numbers.* Princeton University Press.

Champsaur, P., and J.-C. Rochet (1989): "Multiproduct Duopolists," *Econometrica*, pp. 533–557.

Crawford, V. P., and E. M. Knoer (1981): "Job Matching with Heterogeneous Firms and Workers," *Econometrica*, pp. 437–450.

Dupuit, J. (1962): "On Tolls and Transport Charges," *International Economic Papers*, 11.

Eeckhout, J., and P. Kircher (2018): "Assortative Matching with Large Firms," *Econometrica*, 86, 85–132.

Epstein, L., and M. Peters (1999): "A Revelation Principle for Competing Mechanisms," *Journal of Economic theory*, 88, 119–160.

Guerrieri, V., R. Shimer, and R. Wright (2010): "Adverse Selection in Competitive Search Equilibrium," *Econometrica*, 78(6), 1823–1862.

Jullien, B. (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93(1), 1–47.

Kelso, A. S., and V. P. Crawford (1982): "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, pp. 1483–1504.

Laffont, J.-J., and D. Martimort (2002): *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press.

Lester, B., A. Shourideh, V. Venkateswaran, and A. Zetlin-Jones (2018): "Screening and Adverse Selection in Frictional Markets," *Journal of Political Economy*, forthcoming.

Liu, Q., G. J. Mailath, A. Postlewaite, and L. Samuelson (2014): "Stable Matching with Incomplete Information," *Econometrica*, 82(2), 541–587.

Maggi, G., and A. Rodriguez-Clare (1995): "On Countervailing Incentives," *Journal of Economic Theory*, 66(1), 238–263.

Martimort, D., and L. Stole (2002): "The Revelation and Delegation Principles in Common Agency Games," *Econometrica*, 70, 1659–1673.

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

Mussa, M., and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic theory*, 18(2), 301–317.

Rothschild, M., and J. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *The Quarterly Journal of Economics*, pp. 629–649.

Rudin, W. (1987): *Real and Complex Analysis*. McGraw Hill, third edn.

Salop, S. (1979): "Monopolistic Competition with Outside Goods," *The Bell Journal of Economics*, pp. 141–156.

Spulber, D. F. (1989): "Product Variety and Competitive Discounts," *Journal of Economic Theory*, 48(2), 510–525.

Stole, L. A. (1995): "Nonlinear Pricing and Oligopoly," *Journal of Economics & Management Strategy*, 4(4), 529–562.

——— (2007): "Price Discrimination and Competition," *Handbook of industrial organization*, 3, 2221–2299.