Nonlinear pricing with network effects

Arun Sundararajan^{*} Leonard N. Stern School of Business, New York University 44 West 4th Street, KMC 8-93 New York, NY 10012 *asundara@stern.nyu.edu*

While complete and internally consistent, this is a preliminary draft. Feedback is most welcome. If you'd like to quote or circulate any results, please let me know

Abstract

This paper analyzes monopoly pricing for a network good, with heterogeneous customer preferences over both the intrinsic value and the network value of the product. Network effects that depend on gross consumption, individual consumption, and customer type are analyzed. It is shown that network effects generally raise total prices, and may either raise individual consumption for all customer types, lower it for a subset of customer types, or leave it unchanged. These differences highlight the nature of the trade-off between value creation and price discrimination when pricing network goods. The monopolist typically captures all the direct value from the network effects. Customer surplus may increase for those whose individual consumption increases. Consequently, network effects may actually harm low-usage customers, and skew the distribution of surplus towards higher-end users. When the monopolist prices to deter entry, the socially optimal outcome is obtained when network value is constant across customers. On the other hand, when network value is high and depends on individual consumption, the threat of entry has no effect on total welfare, and merely redistributes surplus between the monopolist and the customers. In other cases, optimal consumption increases for a subset of lower-type customers, which mitigates the skew in surplus across customers. Some policy implications of the results are also discussed.

Key Words and Phrases: network externalities, network goods, screening, entry deterrence, information goods

^{*}Draft, January 2003. I thank Roy Radner and Ravi Mantena for helpful discussions, and seminar participants at the University of Pennsylvania and the University of Rochester for feedback on preliminary versions of some of these results. All responsibility for errors, omissions and gratuitous use of Greek letters remains mine.

1 Introduction

This paper analyzes optimal monopoly pricing under incomplete information for a good that displays positive network effects (henceforth called a *network good*). The principal goals of the paper are to characterize the optimal contract and study consumption patterns, profits and welfare, for different kinds of network effects, and under the potential threat of entry.

In standard models of network goods (Katz and Shapiro, 1985, Farrell and Saloner, 1985), it is assumed that each customer purchases a maximum of one unit of the product, that the value of the network effect is proportionate to the total size of the product's eventual user base, and that all customer benefit equally from the network effects. However, there are many network goods which are consumed in variable quantities, and for which the magnitude of the network effect may depend on the total quantity consumed, rather than simply the total number of adopters. In addition, the value each customer get from the network effects may depend on their individual consumption, as well as the intrinsic value they place on the product.

To illustrate the relevance of these observations, consider the purchase of PC operating systems software (a widely used example of a network good) by corporate customers. The simplest pricing problem faced by a seller in this market is one of choosing a pricing schedule, where quantity is measured by number of user licenses, and each corporate customer purchases a variable quantity of licenses. The network effects are caused largely by the higher availability and quality of complementary goods like applications software (word processors, spreadsheet programs, web browsers), as the total number of PC's running the operating system increases. Consequently, the magnitude of the network effects are proportionate to the total number of licenses sold (the gross consumption, rather than simply the number of corporations who adopt the OS). Moreover, a corporation which has a higher number of licenses for the operating system benefits more from the increased quality and availability of the applications – in other words, the value realized from the network effects by any corporate customer also depends on their individual consumption. Finally, even at equal levels of individual consumption, companies who place a higher value on each additional licence (or quantity unit) – for instance, those companies whose employees use their individual copies of the operating system more – are likely to place a higher value on the availability of a wider range of applications. Therefore, the value realized from the network effects by any individual corporate customer may also depend on the customer's type¹.

A similar argument can be made for back-end or enterprise software used in variable quantities by different companies (Oracle's database software, Siebel's CRM solution being two examples),

¹In addition, there is a positive externality driven by value from interoperability, which is far more important within an organization than across companies, and is therefore influenced more by individual consumption.

or for networking equipment such as routers, wherein the network effect is driven by the ease with which one can find qualified support or administration engineers, trained employees, or compatible equipment. Network products and services provided directly to individuals consumers may also display the same properties. For example, electronic marketplaces like eBay are widely recognized as displaying positive network effects, which stem from increased liquidity, as well as a wider availability of robust systems supporting marketplace services such as reputation, escrow, payment, settlement and dispute resolution. The magnitude of the network effects increases not just with the number of participants in the market, but with the extent to which each participant actually buys and sells; moreover, an individual who participates more realizes higher benefits from them. Even for products used as canonical examples of network goods, such as telephone service, usage varies across consumers, network effects dependent on total consumption as well as installed base, users with higher consumption levels benefit more from the network effects, and pricing is often nonlinear, based on consumption, rather than just membership.

The first part of this paper extends the standard models of network goods by characterizing the optimal nonlinear pricing schedule for a monopolist selling a network good whose network effects depend on gross realized consumption, and whose individual customers may realize heterogeneous value from the network effects. Three cases are analyzed successively – where the value to each customer from the network effects depends on just gross consumption, then on both gross consumption and individual consumption, and finally, on gross consumption, individual consumption and customer type. Interestingly, the changes in consumption induced by the network effects varies substantially across the cases – there may be no change in individual consumption for any customer type, an increase in individual consumption across all types, or a reduction in consumption for a fraction of customer types. Examining the welfare properties of the optimal contract in each of these cases (relative to a product with no network effects) suggests that total surplus changes on account of two effects – the direct impact of the value of the network effects. In general, both effects are non-negative; however, the distribution of the surplus between the monopolist and the customers varies across the three cases.

The monopoly model described above is motivated by the fact that markets with positive network effects are often natural monopolies. However, there has been substantial recent interest in whether (and how) the potential *threat of entry* affects the monopoly pricing of network goods. In the recent U.S. versus Microsoft case, both parties agreed that Microsoft's pricing was not consistent with monopoly profit maximization, and Schmalensee (1999) argued that Microsoft underprices in order to reduce the desirability of entry by competing firms into the market for operating systems. Fudenberg and Tirole (2000) develop a formal model of limit pricing that supports this argument, in which installed base plays an entry-deterring role analogous to that of excess capacity (Spence, 1977, Dixit, 1980).

The second part of this paper proposes and analyzes an alternate representation, in which the costless threat of entry results in the monopolist being able to charge each customer no more than the value generated by their product's network effects. In other words, to successfully deter entry, each customer must get surplus equal to at least the maximum intrinsic value they could get from a competing product. This introduces the idea that network value may play the role of being the primary source of profits for a monopolist who prices to successfully deter entry. By construction, in this model, the monopolist cannot extract *any* of the intrinsic value of the product. One would therefore expect the optimal contract to induce a substantial increase in consumption, relative to the model with no entry threat. Surprisingly, it is shown that consumption changes are often zero, that the monopolist may still design contracts that share some of the surplus generated by the network effects, even when the network value is not type-dependent, and that any increases in total surplus are almost always due to consumption increases of a lower subset of types. These optimal contracts are contrasted with those obtained in the absence of the entry threat, and in one case, it is shown that the outcome is socially efficient.

This paper draws from and adds to two lines of research. The first is the literature on the monopoly pricing of network goods (Rohlfs, 1974, Economides, 1996b, Cabral et. al., 1999, Fudenberg and Tirole, 2000, among others). The idea of a fulfilled-expectations equilibrium, which is based on one notion of rational-expectations equilibrium under uncertainty (Radner, 1982), and was introduced into the network externalities literature by Katz and Shapiro (1985) in the context of Cournot competition, is extended to the case of heterogeneous customers purchasing variable quantities. Modeling network goods for which the network effects depend on gross consumption rather than simply network size is new, as is the analysis of heterogeneity in the value of the network effects across customers, based on both individual consumption, and on customer type.

The second line of research is the price screening literature – the specification of the model is a modified version of the standard set-up with a one-dimensional type (Maskin and Riley, 1984), also relies on results from a recent generalization by Jullien (2000) that permits type-dependent participation constraints, and has some outcomes that have a similar structure to duopoly results in Rochet and Stole (2001). The paper contributes new results to the price screening theory by characterizing how positive network effects of different kinds affect optimal nonlinear pricing, and by establishing conditions under which fulfilled-expectations equilibria involving the optimal contracts exist and are unique. It complements recent work by Segal and Whinston (2001), and by Jullien (2001), that examine different problems of optimal contracting in the presence of network externalities. A related line of research is the literature on optimal monopoly pricing with *negative* consumption externalities, specifically in the context of congestion in queuing systems (Mendelson, 1985, Dewan and Mendelson, 1990, Mendelson and Whang, 1990, Westland, 1992).

The rest of this paper is organized as follows. Section 2 specifies the model, and Section 3 illustrates the intuition behind some results in the context of an example with two customer types. Section 4 presents the analysis of the general monopoly model, and Section 5 analyzes the monopolist's problem with the costless threat of entry. Both sections 4 and 5 examine the consumption changes induced by network effects, and some welfare properties of the optimal contracts. Section 6 provides a discussion of the results, the model's assumptions, and concludes with an outline of open research questions.

2 Model

2.1 Firm and customers

A monopolist sells a homogeneous product which may be used by consumers in varying quantities. The creation of the product may have involved expending a fixed cost, which is assumed to be sunk, and has no bearing on the optimal pricing. The variable cost of production is assumed to be zero².

Customers are heterogeneous, indexed by their type $\theta \in [\underline{\theta}, \overline{\theta}]$. The monopolist does not observe the type of any customer, but knows $F(\theta)$, the probability distribution of types in the customer population. $F(\theta)$ is assumed to be strictly increasing and absolutely continuous, and therefore the corresponding density function $f(\theta)$ exists and is strictly positive for all $\theta \in [\underline{\theta}, \overline{\theta}]$. In addition, $\frac{1-F(\theta)}{f(\theta)}$, which is the reciprocal of the hazard rate, is assumed to be non-increasing for all θ . Each customer knows their own type θ .

The preferences of a customer of type θ are represented by the linearly separable utility function

$$V(q,\theta,Q,p) = U(q,\theta) + W(q,\theta,Q) - p,$$
(1)

where q is the quantity of the product used by the customer (often referred to as *individual con*sumption), Q is the total quantity of the product used by all consumers in the market (often referred to as the gross consumption) and p is the total price paid by the customer. The total number of customers in the market is normalized to 1.

For a customer of type θ , the *intrinsic value function* $U(q, \theta)$ represents the intrinsic value of consumption of q units of the product, and the *network value function* $W(q, \theta, Q)$ represents the

²This assumption is made to reflect that fact that a number of network goods are digital. The analysis is largely similar if one introduces a positive and convex cost function c(q). This is discussed more extensively in Section 6.2.

value that the customer gets from the product's network effects, when their individual consumption is q and the gross consumption is Q.

The intrinsic value function $U(q, \theta)$ is assumed to have the following properties:

- 1. $U_{11}(q,\theta) < 0, U_2(q,\theta) > 0, U_{12}(q,\theta) > 0.$ 2. $\frac{d}{d\theta} \left(\frac{-U_{11}(q,\theta)}{U_1(q,\theta)} \right) < 0, U_{122}(q,\theta) \le 0, U_{112}(q,\theta) \ge 0.$
- 3. $\alpha(\theta) = \underset{q}{\arg\max} U(q,\theta)$ is finite and unique for all θ . $U_1(q,\theta) > 0$ for $q < \alpha(\theta)$, and $U_1(q,\theta) < 0$ for $q > \alpha(\theta)$.

Numbered subscripts to functions denote partial derivatives with respect to the corresponding variable. The first set of properties – strict concavity in q, increasing intrinsic value with type, and the Spence-Mirrlees single-crossing condition – are standard. The second set of properties assume decreasing absolute risk aversion (which is standard in models of nonlinear pricing), marginal utility that is concave in type θ (which is a standard assumption that ensures that the optimal contract separates types), and that the concavity of U with respect to q does not increase with type³ (which ensures that the profit function is quasiconcave in quantity).

The third set of properties simply state that there is a consumption level beyond which the intrinsic value of the product decreases. It reflects the reality that customers consume a finite quantity of any product, even if the marginal price of additional consumption is zero. This is because intrinsic value from usage is typically bounded by a constraint on some related resource – attention or computing power being two common examples – and the implicit presence of a substitute use for this resource. Analogously, sometimes the increased consumption of the product may necessitate the purchase of additional costly complementary assets (like computer hardware for software, for instance)⁴.

The network value function $W(q, \theta, Q)$ is assumed to have the following properties:

- 1. $W_1(q, \theta, Q) \ge 0, W_2(q, \theta, Q) \ge 0, W_3(q, \theta, Q) \ge 0, W_{11}(q, \theta, Q) \le 0, W_{12}(q, \theta, Q) \ge 0.$
- 2. $W_{112}(q, \theta, Q) \ge 0, W_{122}(q, \theta, Q) \le 0, W_{123}(q, \theta, Q) \ge 0.$
- 3. $\beta(\theta, Q) = \arg \max_{q} [U(q, \theta) + W(q, \theta, Q)]$ is finite and unique for every θ, Q .

The value from the product's network effects are therefore non-decreasing in individual consumption, type and gross consumption, (weakly) concave in individual consumption, and the marginal

³If $U_1(q,\theta) > 0$, then $U_{112}(q,\theta) \ge 0$ implies that $\frac{d}{d\theta}(\frac{-U_{11}(q,\theta)}{U_1(q,\theta)}) < 0$

⁴See Sundararajan (2002), Section 4, for more discussion and examples. Also Section 6.2 of the current paper discusses relaxing this assumption in the presence of convex costs.

network value $W_1(q, \theta, Q)$ is non-decreasing in type. In addition, even after factoring in the network value, all customers consume a finite quantity of the product even if the marginal price is zero. In all the cases we analyze in detail, W is linear in q, satisfying many of the the conditions above trivially.

The source of the network effects are not modeled explicitly. The model therefore adopts what Economides (1996a) calls the 'macro' approach.

Finally, each customer of type θ is assumed to have reservation utility $\hat{U}(\theta) \ge 0$. The functions $F(\theta), U(q, \theta), W(q, \theta, Q)$, and $\hat{U}(\theta)$ are common knowledge.

2.2 Sequence of events

The interaction between the monopolist and their customers is according to the following sequence:

- 1. The monopolist announces their pricing schedule, which specifies a total payment p(q) for each level of individual consumption q.
- 2. Customers observe p(q), and form an expectation about what the gross consumption under this pricing schedule will be. All customers have access to the same relevant information⁵, and are assumed to form the same expectation Q^E . The monopolist knows what expectation of gross consumption the pricing schedule p(q) will induce.
- 3. Based on their type θ and the expectation of gross consumption Q^E , each customer determines their optimal individual consumption $\tilde{q}(\theta, Q^E) = \arg \max_q U(q, \theta) + W(q, \theta, Q^E) - p(q)$. If the customer gets at least their reservation utility, that is, if:

$$U(\tilde{q}(\theta, Q^E), \theta) + W(\tilde{q}(\theta, Q^E), \theta, Q^E) - p(\tilde{q}(\theta, Q^E)) \ge \hat{U}(\theta),$$
(2)

then the customer chooses to consume $\tilde{q}(\theta, Q^E)$. If not, the customer does not participate, and purchases zero quantity

4. The monopolist gets a payoff of

$$\int_{\boldsymbol{\epsilon} \boldsymbol{\Theta}} p(\tilde{q}(\boldsymbol{\theta}, Q^E)) f(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{3}$$

where $\Theta = \{\theta : U(\tilde{q}(\theta, Q^E), \theta) + W(\tilde{q}(\theta, Q^E), \theta, Q^E) - p(\tilde{q}(\theta, Q^E)) \geq \hat{U}(\theta)\}$ is the set of participating types. Each participating customer gets a payoff of

$$U(\tilde{q}(\theta, Q^E), \theta) + W(\tilde{q}(\theta, Q^E), \theta, Q^A) - p(\tilde{q}(\theta, Q^E)),$$
(4)

⁵The customer's unique knowledge of their own type does not affect their expectation of gross consumption, which is completely determined by $f(\theta)$, p(q) and the functions $U(q, \theta)$, $W(q, \theta, Q)$, and $\hat{U}(\theta)$ (all of which are common knowledge at this stage).

where

$$Q^{A} = \int_{\theta \in \Theta} \tilde{q}(\theta, Q^{E}) f(\theta) d\theta$$
(5)

is the actual realized gross consumption. Each customer that does not participate gets a payoff of $\hat{U}(\theta)$.

2.3 Contracts

The functions p(q) and $\tilde{q}(\theta, Q^E)$ are used merely to illustrate the sequence of interaction precisely, and will not be derived or used in the subsequent analysis. This is because the revelation principle ensures that we can restrict our analysis to *direct mechanisms*. This subsection defines the different contracts, all of which are direct mechanisms, that are used in subsequent analysis.

A simple exposition of mechanism design, the revelation principle and its applications to pricing can be found in chapter 7 of Fudenberg and Tirole (1991). In particular, section 7.2 describes the revelation principle, and section 7.1 discusses a non-linear pricing example.

To simplify notation, the definition of the following contracts is based on the assumption of *full participation* – that is, that all customers find it optimal to purchase under the contract, if the direct mechanism specifies a non-negative allocation for their type. In the models analyzed in sections 4 and 5, full participation is always optimal for the monopolist.

2.3.1 *Q*-feasible contracts

For any expectation of gross consumption Q, a Q-feasible contract is a menu of of quantity-price pairs $q^F(t,Q), \tau^F(t,Q)$ which satisfies incentive-compatibility and individual rationality:

$$[\mathbf{IC}]: \qquad \theta = \arg\max_{t} U(q^{F}(t,Q),\theta) + W(q^{F}(t,Q),\theta,Q) - \tau^{F}(t,Q) \ \forall \theta \tag{6}$$

$$[\mathbf{IR}]: \qquad U(q^F(\theta, Q), \theta) + W(q^F(\theta, Q), \theta, Q) - \tau^F(\theta, Q) \ge \hat{U}(\theta) \ \forall \theta \tag{7}$$

2.3.2 *Q*-optimal contracts

For any expectation of gross consumption Q, an Q-optimal contract $q(\theta, Q), \tau(\theta, Q)$ is a Q-feasible contract that solves the monopolist's profit maximization problem. That is, it solves:

$$\max_{q^F(t,Q),\tau^F(t,Q)} \int_{\underline{\theta}}^{\overline{\theta}} \tau^F(t,Q) f(t) dt,$$
(8)

where $q^F(t, Q)$ and $\tau^F(t, Q)$ are subject to [IC] and [IR].

2.3.3 Feasible fulfilled-expectations contracts

A feasible fulfilled-expectations contract is a menu of price-quantity pairs $q^{FE}(\theta), \tau^{FE}(\theta)$ such that the contract $q^F(t, Q), \tau^F(t, Q)$ defined by

$$Q = \int_{\underline{\theta}}^{\theta} q^{FE}(\theta) f(\theta) d\theta$$

$$q^{F}(t,Q) = q^{FE}(t) \qquad (9)$$

$$\tau^{F}(t,Q) = \tau^{FE}(t)$$

is a Q-feasible contract. As a consequence, if any Q-feasible contract $q^F(t,Q), \tau^F(t,Q)$ satisfies fulfilled-expectations:

$$[\mathbf{FE}]: Q = \int_{\underline{\theta}}^{\overline{\theta}} q^F(t, Q) f(t) dt, \qquad (10)$$

then the contract $q^{FE}(\theta) = q^F(\theta, Q), \tau^{FE}(\theta) = \tau^F(\theta, Q)$ is a feasible fulfilled-expectations contract.

2.3.4 Optimal fulfilled-expectations contracts

A optimal fulfilled-expectations contract is a menu of price-quantity pairs $q^*(\theta), \tau^*(\theta)$ such that the contract $q(\theta, Q), \tau(\theta, Q)$ defined by

$$Q = \int_{\underline{\theta}}^{\overline{\theta}} q(\theta) f(\theta) d\theta$$

$$q(\theta, Q) = q^{*}(\theta)$$

$$\tau(\theta, Q) = \tau^{*}(\theta)$$
(11)

is an Q-optimal contract. Therefore, if any Q-optimal contract $q(\theta, Q), \tau(\theta, Q)$ satisfies fulfilledexpectations:

$$[\mathbf{FE}]: Q = \int_{\underline{\theta}}^{\overline{\theta}} q(\theta, Q) f(\theta) d\theta, \qquad (12)$$

then the contract $q^*(\theta) = q(\theta, Q), \tau^*(\theta) = \tau(\theta, Q)$ is a fulfilled-expectations optimal contract.

We assume that the solution to the monopolist's pricing problem is always a optimal fulfilledexpectations contract. The conditions for the existence and possible uniqueness of these contracts are described independently in each subsection.

2.4 Direct and indirect surplus changes

Changes in total surplus as a consequence of the network effects can characterized by benchmarking the surplus derived in the model against a *base case* scenario where there are no network effects. This subsection characterizes the base case, and defines direct and indirect surplus changes, which are examined in some of the models in Sections 3 and 4.

Define $q^0(\theta), \tau^0(\theta)$ to be the profit maximizing contract offered by the monopolist under the base case, that is, when $W(q, \theta, Q) = 0$. In all the models analyzed subsequently, this contract is unique. The gross consumption under the base case is

$$Q^{0} = \int_{\underline{\theta}}^{\overline{\theta}} q^{0}(\theta) f(\theta) d\theta$$
(13)

Suppose $q^*(\theta), \tau^*(\theta)$ is an optimal fulfilled-expectations contract for some network value function $W(q, \theta, Q)$, and let

$$Q^* = \int_{\underline{\theta}}^{\overline{\theta}} q^*(\theta) f(\theta) d\theta.$$
(14)

Relative to the base case, the net change in total surplus as a consequence of the network effects is therefore:

$$\int_{\underline{\theta}}^{\overline{\theta}} [U(q^*(\theta), \theta) + W(q^*(\theta), \theta, Q^*)] f(\theta) d\theta - \int_{\underline{\theta}}^{\overline{\theta}} U(q^0(\theta), \theta) f(\theta) d\theta.$$
(15)

The *direct change in surplus* from a customer of type θ as a consequence of the network effects is defined as:

$$s^{n}(\theta) = W(q^{0}(\theta), \theta, Q^{0}), \tag{16}$$

and the *indirect change in surplus* from a customer of type θ as a consequence of the network effects is defined as

$$s^{q}(\theta) = [U(q^{*}(\theta), \theta) - U(q^{0}(\theta), \theta)] + [W(q^{*}(\theta), \theta, Q^{*}) - W(q^{0}(\theta), \theta, Q^{0})].$$
(17)

 $s^{n}(\theta)$ measures the direct change in surplus as a consequence of having the *increase in value* from the network effects, without accounting for any of the changes in consumption. $s^{q}(\theta)$ measures the changes in surplus that arise indirectly as a consequence of the changes in *consumption* that the network effects induce. The total change in surplus across all types, as specified in (15), can now be equivalently expressed as $\int_{\theta}^{\overline{\theta}} [s^{n}(\theta) + s^{q}(\theta)]f(\theta)d\theta$.

3 Example: Two customer types

To illustrate some of the analysis to follow, consider the case where customers are of two types θ_L and θ_H , with $\theta_L < \theta_H$. The probability that a customer is of type θ_H is denoted λ , and the probability that the customer is of type θ_L is $(1 - \lambda)$. Since there are just two types, we drop the functional notation for contracts, expressing then as pairs of price-quantity pairs, for which $q_j^i = q^i(\theta_j)$ and $\tau_j^i = \tau^i(\theta_j)$.

The monopolist's problem is to find the optimal fulfilled-expectations contract $(q_H^*, \tau_H^*), (q_L^*, \tau_L^*)$. In this section, we assume that $\hat{U}(\theta_H) = \hat{U}(\theta_L) = 0$.

Under the base case, it is straightforward to establish that the monopolist's profit maximizing contract $(q_H^0, \tau_H^0), (q_L^0, \tau_L^0)$ satisfies:

$$\frac{U_1(q_L^0, \theta_L)}{U_1(q_L^0, \theta_H)} = \lambda, \tag{18a}$$

$$U_1(q_H^0, \theta_H) = 0, \tag{18b}$$

and

$$\tau_L^0 = U(q_L^0, \theta_L), \tag{18c}$$

$$\tau_H^0 = U(q_H^0, \theta_H) - [U(q_L^0, \theta_H) - U(q_L^0, \theta_L)].$$
(18d)

3.1 *Q*-proportional network effects

We start with the simplest kind of network effects, for which $W(q, \theta, Q) = w(Q)$. It is well-known that in the two-type model with equal reservation utilities, a contract which satisfies [IC] for the higher type, and [IR] for the lower type, is incentive-compatible and individually rational for both types. Given an expectation of gross consumption Q, any Q-feasible contract $(q_H^F, \tau_H^F), (q_L^F, \tau_L^F)$ therefore satisfies:

$$[IC] : U(q_H^F, \theta_H) + w(Q) - \tau_H^F \ge U(q_L^F, \theta_H) + w(Q) - \tau_L^F$$
(19a)

$$[IR] : U(q_L^F, \theta_L) + w(Q) - \tau_L^F \ge 0.$$
(19b)

The monopolist maximizes expected profits $\lambda \tau_H^F + (1-\lambda)\tau_L^F$. Any *Q*-optimal contract would satisfy (19a) and (19b) as equalities. Solving these for the τ values, and maximizing expected profits yields the conditions any *Q*-optimal contract $(q_H, \tau_H), (q_L, \tau_L)$ must satisfy:

$$\tau_L = U(q_L, \theta_L) + w(Q) \tag{20a}$$

$$\tau_H = U(q_H, \theta_H) - U(q_L, \theta_H) + \tau_L$$
(20b)

$$U_1(q_L, \theta_L) = \lambda U_1(q_L, \theta_H)$$
(20c)

$$U_1(q_H, \theta_H) = 0. \tag{20d}$$



Figure 1: Illustrates the optimal allocation and pricing with two consumer types and with network value $W(q, \theta, Q) = w(Q)$, which depends on just gross consumption.

Therefore, the optimal fulfilled-expectations contract $(q_H^*, \tau_H^*), (q_L^*, \tau_L^*)$ satisfies:

$$\frac{U_1(q_L^*, \theta_L)}{U_1(q_L^*, \theta_H)} = \lambda;$$
(21a)

$$U_1(q_H^*, \theta_H) = 0; \tag{21b}$$

and

$$\tau_L^* = U(q_L^*, \theta_L) + w(\lambda q_H^* + (1 - \lambda) q_L^*);$$
(21c)

$$\tau_{H}^{*} = U(q_{H}^{*}, \theta_{H}) - U(q_{L}^{*}, \theta_{H}) + U(q_{L}^{*}, \theta_{L}) + w(\lambda q_{H}^{*} + (1 - \lambda)q_{L}^{*}).$$
(21d)

It can be verified that the solution is unique for any strictly concave $U(q, \theta)$.

The optimal fulfilled-expectations contract is illustrated in Figure 1. Relative to base case, the customer utility functions are shifted up by $w(Q^*)$ for both types, at all levels of consumption. This results in optimal quantities that are identical to those under the base case.

Correspondingly, the optimal prices for both types increase by exactly $w(Q^*)$ relative to the base case. The direct increase in surplus is appropriated entirely by the monopolist. Since consumption is unaltered, there are no indirect changes in surplus. The average price per unit increases for both types, and this increase is more for the lower type.

3.2 q, Q-proportional network effects

The next example involves network effects that are proportionate to both gross consumption and individual consumption. The network value function is assumed to take the linear form $W(q, \theta, Q) = qw(Q)$. Therefore, all *Q*-feasible contracts $(q_H^F, \tau_H^F), (q_L^F, \tau_L^F)$ satisfy:

$$U(q_H^F, \theta_H) + q_H^F w(Q) - \tau_H^F \geq U(q_L^F, \theta_H) + q_L^F w(Q) - \tau_L^F$$
(22a)

$$U(q_L^F, \theta_L) + q_L^F w(Q) - \tau_L^F \ge 0.$$
(22b)

Maximizing monopoly profits yields the following expressions for any Q-optimal contract $(q_H, \tau_H), (q_L, \tau_L)$:

$$\tau_L = U(q_L, \theta_L) + q_L w(Q) \tag{23a}$$

$$\tau_H = U(q_H, \theta_H) - U(q_L, \theta_H) + U(q_L, \theta_L) + q_H w(Q)$$
(23b)

$$U_1(q_L, \theta_L) + w(Q) = \lambda(U_1(q_L, \theta_H) + w(Q))$$
(23c)

$$U_1(q_H, \theta_H) + w(Q) = 0 \tag{23d}$$

Therefore, the optimal fulfilled-expectations contract $(q_H^*, \tau_H^*), (q_L^*, \tau_L^*)$ satisfies:

$$\frac{U_1(q_L^*, \theta_L) + w(Q^*)}{U_1(q_L^*, \theta_H) + w(Q^*)} = \lambda,$$
(24a)

$$U_1(q_H, \theta_H) + w(Q^*) = 0,$$
 (24b)

and

$$\tau_L^* = U(q_L^*, \theta_L) + q_L^* w(Q^*), \tag{24c}$$

$$\tau_H^* = U(q_H^*, \theta_H) + q_H^* w(Q^*) - [U(q_L^*, \theta_H) - U(q_L^*, \theta_L)],$$
(24d)

where $Q^* = \lambda q_H^* + (1 - \lambda) q_L^*$.

The optimal fulfilled-expectations contract for each type is illustrated in Figure 2. The network value shifts the customer utility functions up by $qw(Q^*)$ for both types. Since the shift is proportionate to quantity, this results in optimal quantities that are different from those of the base case. Comparing (24a) and (24b) with (21a) and (21b), it can be shown that both q_H^* and q_L^* are higher than the corresponding quantity levels in the base case. Therefore, in addition to the surplus increase from the network value, there is also an indirect surplus change as a consequence of the increase in consumption across both types. Correspondingly, total prices paid by both types also increase.



Figure 2: Illustrates the optimal allocation and pricing with two consumer types and with network value $W(q, \theta, Q) = qw(Q)$, which depends on individual consumption and gross consumption.

To establish that the monopolist appropriates all of the direct increase in surplus, consider the following *feasible* fulfilled-expectations contract:

$$q_L^{FE} = q_L^0, \ \tau_L^{FE} = U(q_L^0, \theta_L) + q_L^0 w(Q^0);$$
(25a)

$$q_H^{FE} = q_H^0, \ \tau_H^{FE} = U(q_H^0, \theta_H) + q_H^0 w(Q^0) - [U(q_L^0, \theta_H) - U(q_L^0, \theta_L)].$$
(25b)

Relative to the base case, this contract provides the monopolist with a profit increase of $\lambda q_L^0 w(Q^0) + (1-\lambda)q_H^0 w(Q^0)$. Since $s^q(\theta_i) = q_i^0 w(Q^0)$, this profit increase is *exactly* equal to the expected *direct increase* in surplus across both types under the optimal fulfilled-expectations contract $(q_H^*, \tau_H^*), (q_L^*, \tau_L^*)$. Clearly, the optimal fulfilled-expectations contract results in a (weakly) higher increase in surplus for the monopolist.

In addition, the expression in square brackets in (24d), which represents the surplus of the higher type customer, is determined by q_L^* , and $q_L^* > q_L^0$. Therefore, the surplus appropriated by the higher type customer also increases. We can therefore conclude that the monopolist appropriates all of the direct increase in surplus, and that the monopolist and the higher type customer share the indirect increase in surplus. As is customary, the lower type customer gets no surplus. This highlights an important trade-off faced by the monopolist when considering inducing higher consumption from low-end customers. If the monopolist were to subsidize the lower type customer further, it would increase gross consumption, and consequently raise the increase in surplus from the network effects (and the monopolist's profit potential). However, this increase in consumption would also reduces the monopolist's ability to extract the indirect increases in surplus from highertype customers. This issue is discussed further in subsequent sections.

3.3 q, θ, Q -proportional network effects

Finally, consider an example in which network effects are proportionate to gross consumption, individual consumption and type. The network value function is assumed to take the linear form $W(q, \theta, Q) = qw(Q, \theta)$. All Q-feasible contracts $(q_H^F, \tau_H^F), (q_L^F, \tau_L^F)$ therefore satisfy:

$$U(q_H^F, \theta_H) + q_H^F w(Q, \theta_H) - \tau_H^F \geq U(q_L^F, \theta_H) + q_L^F w(Q, \theta_H) - \tau_L^F$$
(26a)

$$U(q_L^F, \theta_L) + q_L^F w(Q, \theta_L) - \tau_L^F \ge 0.$$
(26b)

Proceeding as before, given Q, the conditions any Q-optimal contract $(q_H, \tau_H), (q_L, \tau_L)$ must satisfy are:

$$\tau_L = U(q_L, \theta_L) + q_L w(Q, \theta_L)$$
(27a)

$$\tau_H = U(q_H, \theta_H) - U(q_L, \theta_H) + U(q_L, \theta_L) + (q_H - q_L)w(Q)$$
 (27b)

$$U_1(q_L, \theta_L) + w(Q, \theta_L) = \lambda(U_1(q_L, \theta_H) + w(Q, \theta_H))$$
(27c)

$$U_1(q_H, \theta_H) + w(Q, \theta_H) = 0 \tag{27d}$$

The unique optimal fulfilled-expectations contract therefore satisfies:

$$\frac{U_1(q_L^*, \theta_L) + w(Q^*, \theta_L)}{U_1(q_L^*, \theta_H) + w(Q^*, \theta_H)} = \lambda,$$
(28a)

$$U_1(q_H, \theta_H) + w(Q^*, \theta_H) = 0,$$
 (28b)

and

$$\tau_L^* = U(q_L^*, \theta_L) + q_L^* w(Q^*, \theta_L),$$
(28c)

$$\tau_{H}^{*} = U(q_{H}^{*}, \theta_{H}) + q_{H}^{*}w(Q^{*}, \theta_{H})$$
(28d)

$$-[U(q_L^*, \theta_H) - U(q_L^*, \theta_L)] - q_L^*[w(Q^*, \theta_H) - w(Q^*, \theta_L)],$$

where $Q^* = \lambda q_H^* + (1 - \lambda) q_L^*$.

The optimal contract for each type is illustrated in Figure 3. The network value shifts the customer utility functions up for both types, but more for the higher type, and more for higher



Figure 3: Illustrates the optimal allocation and pricing with two consumer types and with network value $W(q, \theta, Q) = qw(Q, \theta)$, which depends on type, individual consumption and gross consumption.

consumption levels. It is clear that this results in an increase in consumption for the higher type customer. However, the consumption of the lower type can either increase or decrease.

One can establish that if $\frac{U(q_L^*, \theta_L)}{U(q_L^*, \theta_H)} < (>) \frac{w(Q^*, \theta_L)}{w(Q^*, \theta_H)}$, then $q_L^* > (<) q_L^0$. This is based on the fact that $\frac{U(q, \theta_L)}{U(q, \theta_H)}$ is strictly decreasing in q, and the equalities in (18a) and (28a). That is, under the optimal fulfilled-expectations contract, if the marginal benefit from the network effects are substantially higher for higher type customers, the monopolist *reduces* consumption for lower types. The intuition behind this observation is clearer if one examines the negative expressions in square brackets at the end of (28d), which represent the surplus captured by the higher type. Similar to section 3.2, the surplus captured by the higher type increases with q_L^* . However, as q_L^* increases, the marginal increase in the surplus the monopolist must share with the customer is higher at higher values of $[w(Q^*, \theta_H) - w(Q^*, \theta_L)]$, and this provides the incentive for the monopolist to reduce q_L^* below q_L^0 .

4 Monopoly with network effects

We now return to the general model of monopoly with network effects specified in Section 2, which is solved in this section. The analysis in this section makes the following assumption:

$$\hat{U}(\theta) = 0 \text{ for all } \theta.$$
⁽²⁹⁾

This assumption is relaxed in Section 5, when there is a threat of entry.

4.1 Base case: no network effects

The purpose of this subsection is to characterize the optimal contract offered by the monopolist in the absence of network effects (the base case, as discussed in section 2.4). Since there are no network effects, fulfilled-expectations do not play a role.

Lemma 1 When $W(q, \theta, Q) = 0$, the monopolist offers the contract $q^0(\theta), \tau^0(\theta)$ which satisfies the following conditions for all θ :

$$\frac{U_1(q^0(\theta),\theta)}{U_{12}(q^0(\theta),\theta)} = \frac{1-F(\theta)}{f(\theta)};$$
(30)

$$\tau^{0}(\theta) = U(q^{0}(\theta), \theta) - \int_{\underline{\theta}}^{\theta} U_{2}(q^{0}(x), x) dx$$
(31)

This contract defined by (30) and (31) is unique. Moreover, for all θ such that $q^0(\theta) > 0$, it satisfies $q_1^0(\theta) > 0$, $\tau_1^0(\theta) > 0$.

The proof of this result is omitted. The reader is referred to chapter 2 of Salanié (1997) for a simple exposition, or to Maskin and Riley (1984) for more details. A complete proof based on a model formulation similar to that of this paper is also available in Sundararajan (2002).

4.2 Structure of optimal contracts

This subsection describes the structure of Q-optimal contracts, and demonstrates their uniqueness. It also provides the necessary conditions for an optimal fulfilled-expectations contract. Subsequent sections, which impose a more specific structure on $W(q, \theta, Q)$ provide sufficient conditions for existence and uniqueness.

Lemma 2 For every expectation of consumption Q, the Q-optimal contract $q(\theta, Q), \tau(\theta, Q)$ is unique, and is defined by the following conditions:

$$\frac{U_1(q(\theta, Q), \theta) + W_1(q(\theta, Q), \theta, Q)}{[U_{12}(q(\theta, Q), \theta) + W_{12}(q(\theta, Q), \theta, Q)]} = \frac{1 - F(\theta)}{f(\theta)},$$
(32)

and

$$\tau(\theta, Q) = U(q(\theta, Q), \theta) + W(q(\theta, Q), \theta, Q) - \int_{\underline{\theta}}^{\theta} [U_2(q(\theta, Q), x) + W_2(q(\theta, Q), x, Q)] dx.$$
(33)

Unless otherwise specified, proofs of all results are available in Appendix A. The uniqueness of the *Q*-optimal contract is important, especially in the analysis of entry-deterrence in Section 5.2. Based on Lemma 2, characterizing an optimal fulfilled-expectations equilibrium is straightforward.

Lemma 3 Any optimal fulfilled-expectations contract $q^*(\theta), \tau^*(\theta)$ satisfies the following conditions:

$$Q^* = \int_{\underline{\theta}}^{\overline{\theta}} q^*(\theta) f(\theta) d\theta, \qquad (34)$$

$$\frac{U_1(q^*(\theta), \theta) + W_1(q^*(\theta), \theta, Q^*)}{[U_{12}(q^*(\theta), \theta) + W_{12}(q^*(\theta), \theta, Q^*)]} = \frac{1 - F(\theta)}{f(\theta)},$$
(35)

and

$$\tau^{*}(\theta) = U(q^{*}(\theta), \theta) + W(q^{*}(\theta), \theta, Q^{*}) - \int_{\underline{\theta}}^{\theta} [U_{2}(q^{*}(x), x) + W_{2}(q^{*}(x), x, Q^{*})] dx.$$
(36)

Lemma 3 follows directly from the first-order necessary conditions for the monopolist's problem, reformulated in the standard way, and with the additional requirement of fulfilled expectations.

4.3 *Q*-proportional network effects

This subsection specifies the optimal fulfilled-expectations contract when network value function depends on just gross consumption, and discusses some properties of consumption, pricing and welfare under this contract.

The following proposition establishes that the unique solution to the monopolist's problem in this case is very similar to that of the base case:

Proposition 1 If $W(q, \theta, Q) = w(Q)$, then the optimal fulfilled-expectations contract takes the form:

$$q^*(\theta) = q^0(\theta); \tag{37}$$

$$\tau^*(\theta) = \tau^0(\theta) + w(Q^0),$$
 (38)

where $q^{0}(\theta)$ and $\tau^{0}(\theta)$ are specified in (30) and (31), and $Q^{0} = \int_{\underline{\theta}}^{\overline{\theta}} q^{0}(\theta) f(\theta) d\theta$. A contract of this form exists and is unique for any network value function w(Q).



Figure 4: Based on Proposition 1, illustrates the optimal consumption of two types θ_1 and θ_2 ($\theta_1 < \theta_2$) when network effects are proportionate to just Q. The point q at which first-order necessary conditions are met for each type is at the intersection of the $U_1(q,\theta)$ and the $U_{12}(q,\theta)\frac{1-F(\theta)}{f(\theta)}$ curves, neither of which is altered by a network value function dependent on just Q. As a consequence, $q^*(\theta) = q^0(\theta)$.

Proposition 1 shows that when the network value function depends on just gross consumption, the monopolist finds it optimal to induce levels of consumption from each customer type that are identical to those in the absence of network effects, and simply increase the price charged to every type by an amount equal to the network value. The form of this result is similar to that obtained in the two-type example of section 3.1 – consumption remains unaltered, and total price goes up equally for all types.

The intuition behind this result is straightforward. For any common expectation Q of gross consumption, all of the customers' utility functions are shifted up by the same constant amount, equal to w(Q). Since there is no change in the marginal properties of the utility functions, the monopolist's optimal allocation remains the same. This is illustrated in Figure 4.

It is evident from (38) that the monopolist captures all of the direct increase in surplus. Since there is no change in consumption relative to the base case, there is no indirect change in surplus. Customer surplus is therefore identical to the base case, for all types.

4.4 q, Q-proportional network effects

This subsection characterizes the optimal fulfilled-expectations contract when the network value function depends on both gross consumption and individual consumption, specifies the conditions under which a contract of this form exists, and is unique, and also discusses changes in consumption, pricing and welfare, relative to the base case.

The following proposition establishes the main results of the subsection:

Proposition 2 (a) If $W(q, \theta, Q) = qw(Q)$, then any optimal fulfilled-expectations contract satisfies the following conditions:

$$\frac{U_1(q^*(\theta), \theta) + w(Q^*)}{U_{12}(q^*(\theta), \theta)} = \frac{1 - F(\theta)}{f(\theta)},$$
(39)

and

$$\tau^*(\theta) = U(q^*(\theta), \theta) + q^*(\theta)w(Q^*) - \int_{\underline{\theta}}^{\theta} U_2(q^*(x), x)dx,$$
(40)

where $Q^* = \int_{\underline{\theta}}^{\overline{\theta}} q^*(\theta) f(\theta) d\theta$.

(b) If w(Q) has a finite upper bound \overline{w} , then an optimal fulfilled-expectations contract always exists. In addition, for all Q, if $w_1(Q) \leq -U_{11}(q,\overline{\theta})$ for all Q and q, then (39) and (40) specify the unique optimal fulfilled-expectations contract.

(c) For all θ , $q^*(\theta) > q^0(\theta)$, and $\tau^*(\theta) > \tau^0(\theta)$.

(d) The monopolist extracts all of the direct increase in surplus, and shares some of the indirect increase in surplus with the customers. That is:

$$\int_{\underline{\theta}}^{\overline{\theta}} \tau^*(\theta) f(\theta) d\theta - \int_{\underline{\theta}}^{\overline{\theta}} \tau^0(\theta) f(\theta) d\theta > \int_{\underline{\theta}}^{\overline{\theta}} s^n(\theta) f(\theta) d\theta,$$
(41)

and

$$\int_{\underline{\theta}}^{\overline{\theta}} \tau^*(\theta) f(\theta) d\theta - \int_{\underline{\theta}}^{\overline{\theta}} \tau^0(\theta) f(\theta) d\theta < \int_{\underline{\theta}}^{\overline{\theta}} [s^n(\theta) + s^q(\theta)] f(\theta) d\theta,$$
(42)

where $s^{n}(\theta)$ and $s^{q}(\theta)$ are as defined in (16) and (17).

Sufficient conditions for the existence of an optimal fulfilled-expectations equilibrium are fairly mild – all that is required is that the marginal benefit from the network effects w(Q) be bounded. For a unit change in Q, the condition for uniqueness restricts the marginal network value to being less than the marginal increase in intrinsic value for the highest type. In general, this requires that network value not grow too fast relative to intrinsic value. The proof of proposition 2 specifies



Figure 5: Based on Proposition 2, illustrates the optimal consumption of two types θ_1 and θ_2 $(\theta_1 < \theta_2)$ when network effects are proportionate to q and Q. The marginal intrinsic value curves $U_1(q, \theta)$ are shifted up by a constant amount w(Q), while the $U_{12}(q, \theta) \frac{1-F(\theta)}{f(\theta)}$ curves remain the same as in the base case. Since the former set of curves slope down, and the latter slope up, this results in a strict increase in consumption for all types, relative to the base case.

stronger alternate sufficient conditions, which are independent of w(Q), but place restrictions on the curvature of $U(q, \theta)$ that depend on the reciprocal of the hazard rate, and may not always be met.

Even if the solution is not unique, this is not unduly troubling, since multiple possible equilibrium outcomes are not uncommon in models of network goods. The monopolist simply needs to pick the optimal fulfilled-expectations contract that provides the highest profits – customer expectations are formed after the contract is specified, anyway. It is more important to note that the results in parts (c) and (d) of the proposition *do not* rely on uniqueness. Therefore, when multiple optimal fulfilled-expectations contracts exist, each of them will result in higher individual consumption and price, and each will induce the surplus distribution specified in (41) and (42).

The result from part (c) is illustrated in Figure 5, for two candidate types. As in section 4.3, the result is consistent with the one obtained from the two-type model⁶. Part (d) establishes that the monopolist continues to get all the direct increase in surplus from the network effects, and that

⁶It is also likely that the optimal consumption of the higher types increases more than that of the lower types, though establishing this rigorously requires a number of detailed assumptions on the exact curvature of $U(q, \theta)$.

any increase in surplus for customers comes from the indirect increases. This extends the result in section 4.3, where there was no indirect increase in surplus. In addition, it is consistent with the two-type model in section 3.2. The result follows from a similar argument – that the feasible fulfilled-expectations contract that implements $q^0(\theta)$ results in a profit increase equal to the direct increase in surplus.

4.5 q, θ, Q -proportional network effects

This subsection characterizes the optimal fulfilled-expectations contract when the network value function depends on gross consumption and individual consumption, as well as on customer type, specifies sufficient conditions under which a contract of this form exists and is unique, and discusses changes in consumption and pricing.

Proposition 3 characterizes the optimal fulfilled-expectations contract, sufficient conditions for its existence and uniqueness, and the changes in individual consumption for each type relative to the base case.

Proposition 3 (a) If $W(q, \theta, Q) = qw(Q, \theta)$, then any optimal fulfilled-expectations contract satisfies the following conditions:

$$\frac{U_1(q^*(\theta), \theta) + w(Q^*, \theta)}{U_{12}(q^*(\theta), \theta) + w_2(Q^*, \theta)} = \frac{1 - F(\theta)}{f(\theta)},\tag{43}$$

and

$$\tau^{*}(\theta) = U(q^{*}(\theta), \theta) + q^{*}(\theta)w(Q^{*}) - \int_{\underline{\theta}}^{\theta} [U_{2}(q^{*}(x), x) + q^{*}(x)w_{2}(Q^{*}, x)]dx,$$
(44)

where $Q^* = \int_{\underline{\theta}}^{\overline{\theta}} q^*(\theta) f(\theta) d\theta$.

(b) If $w(Q,\overline{\theta})$ has a finite upper bound \overline{w} , then an optimal fulfilled-expectations contract always exists. In addition, for all Q, if $w_1(Q,\overline{\theta}) \leq -U_{11}(q,\overline{\theta})$ for all Q and q, then (43) and (44) specify the unique optimal fulfilled-expectations contract.

(c) For each θ , if $\frac{U_1(q^*(\theta), \theta)}{U_{12}(q^*(\theta), \theta)} > (<) \frac{w(Q^*, \theta)}{w_2(Q^*, \theta)}$, then $q^*(\theta) < (>) q^0(\theta)$. However, there always exists a subset of types $[\hat{\theta}, \overline{\theta}]$ such that $q^*(\theta) \ge q^0(\theta)$ for all $\theta \in [\hat{\theta}, \overline{\theta}]$.

Proposition 3 indicates that an upper bound on the marginal network value is sufficient to ensure that an optimal fulfilled-expectations contract exists, and that limiting the rate of change of marginal network value ensures uniqueness. Again, alternative sufficient conditions for uniqueness are discussed in the proof of Proposition 3. It is possible that in many applications of the model,



Figure 6: Based on Propositions 3 and 4, illustrates the optimal consumption of two types θ_1 and θ_2 ($\theta_1 < \theta_2$) when network effects are proportionate to q, θ and Q. The marginal intrinsic value curves $U_1(q,\theta)$ are shifted up by different amounts $w(Q,\theta)$, while the $U_{12}(q,\theta)\frac{1-F(\theta)}{f(\theta)}$ curves are shifted up by $w_2(Q,\theta)$. In the figure, the marginal network value $w(Q,\theta)$ increases from θ_1 to θ_2 , at a relatively steady value of $w_2(Q,\theta)$. The increased heterogeneity between types results in a reduction in consumption for θ_1 when the monopolist optimally price-discriminates.

there may be multiple optimal fulfilled-expectations contracts; however, the results of part (c) and of Proposition 4 apply to *any* optimal fulfilled-expectations contract.

Surprisingly, part (c) of the proposition shows that individual consumption may actually reduce for some customer types, despite the fact that network effects cause both the total utility and the marginal utility to increase for all customers, at any consumption level. The result also indicates that the set of customers for whom consumption reduces are not part of a highest subset of types. This reduction in consumption is caused by optimal price-discrimination when network value increases faster than intrinsic value. At comparable levels of marginal intrinsic value $U_1(q, \theta)$ and marginal network value $w(Q, \theta)$, when the rate at which the marginal network value increases with type is substantially higher than the rate at which intrinsic value increases with type – that is, when $w_2(Q^*, \theta) > U_{12}(q^*(\theta), \theta)$ – it becomes difficult for the monopolist to ensure incentive compatibility for the higher types, necessitating a reduction in the price and consumption levels of lower types. The result accentuates the trade-off between price discrimination and increasing consumption that was initially highlighted at the end of section 3.2.

Part (c) of Proposition 3 also shows that there is always a subset of types for whom consumption increases. Under some fairly weak additional assumptions, further structure can be placed on the changes in individual consumption, relative to the base case:

Proposition 4 Suppose the marginal network value $w(Q, \theta)$ is concave in θ . In addition, assume that either $\frac{\partial}{\partial \theta} \left(\frac{U_1(q,\theta)}{U_{12}(q,\theta)} \right) \leq \frac{1+U_{12}(q,\theta)}{U_{12}(q,\theta)}$, or that $U_{112}(q,\theta) \geq U_{11}(q,\theta)U_{122}(q,\theta)$. Then: (a) If $\frac{U_1(q^*(\underline{\theta}),\underline{\theta})}{U_{12}(q^*(\underline{\theta}),\underline{\theta})} < \frac{w(Q^*,\underline{\theta})}{w_2(Q^*,\underline{\theta})}$, then $q^*(\theta) > q^0(\theta)$ for all $\theta \in [\underline{\theta},\overline{\theta}]$ (b) If $\frac{U_1(q^*(\underline{\theta}),\underline{\theta})}{U_{12}(q^*(\underline{\theta}),\underline{\theta})} > \frac{w(Q^*,\underline{\theta})}{w_2(Q^*,\underline{\theta})}$, then there exists an interior type $\hat{\theta} \in (\underline{\theta},\overline{\theta})$ such that $q^*(\theta) < q^0(\theta)$ for $\theta < \hat{\theta}$, and $q^*(\theta) > q^0(\theta)$ for $\theta > \hat{\theta}$.

Proposition 4 establishes that if the consumption $q^*(\underline{\theta})$ of the lowest type increases under the new contract, then so does the consumption of all types. More generally, if $q^*(\theta) \ge q^0(\theta)$ for any type θ , then $q^*(\theta) > q^0(\theta)$ for all higher types. Since $w(Q^*, \overline{\theta}) > 0$, part (a) of Proposition 3 ensures that the consumption of the highest type is always higher than $q^0(\overline{\theta})$, and that, in fact, $U_1(q^*(\overline{\theta}), \overline{\theta}) < 0$. There is therefore always a positive subset of customer types for whom consumption increases.

Either of the additional assumptions ensures monotonicity in the difference between the rates at which $\frac{U_1(q,\theta)}{U_{12}(q,\theta)}$ and $\frac{w(Q,\theta)}{w(Q,\theta)}$ vary with θ at the optimal contract $q^*(\theta)$, which enables the strict ordering of types that the proposition establishes. These assumptions are not 'tight', and it is possible that they are not necessary, but establishing this remains work-in-progress.

It is straightforward to establish that relative to the base-case, there is a strict increase in the monopolist's profits, and that there is a strict increase in surplus for any customer whose consumption $q^*(\theta) > q^0(\theta)$. Since there is always a positive measure of such types, the total increase in surplus from the network effects (direct and indirect) is never fully appropriated by either the seller or the customers. Beyond that, however, it is harder to characterize the welfare properties of the contract.

5 Entry-deterring monopoly with network effects

This section analyzes changes in the monopolist's optimal pricing strategy when facing a threat of entry from a single entrant, whose network good is intrinsically a perfect substitute for the monopolist's product. By virtue of being the incumbent, the monopolist's product has a positive network value for all customers. The entrant's product, on the other hand, provides only its intrinsic value to the customers. The fixed cost of entry is assumed to be zero.

5.1 Formulation of the monopolist's problem

The purpose of this subsection is to establish that the problem of pricing to deter entry under the threat of costless entry is equivalent to a problem of pricing in the absence of any entry threat, but with type-dependent individual rationality constraints.

The utility of a customer of type θ who purchases a quantity q of the monopolist's product for a payment p is

$$U(q,\theta) + W(q,\theta,Q) - p, \tag{45}$$

and the utility of a customer of type θ who purchases a quantity q of the entrant's product for a payment p is

$$U(q,\theta) - p. \tag{46}$$

Given a set of prices, and an expectation Q of gross consumption of the monopolist's product, customers choose the product and quantity that maximizes their utility. Customers indifferent between the monopolist's and the entrant's products are assumed to choose the monopolist's product.

A complete characterization of the entry game is not provided. Rather, the analysis focuses on the characteristics of pricing schedules for the monopolist that successfully *deter entry*. Since the fixed cost of entry is assumed to be zero, these are pricing schedules under which any pricing scheme offered by the entrant results in zero profits for the entrant.

Recall that

$$\alpha(\theta) = \arg\max_{q} U(q, \theta), \tag{47}$$

and that

$$\beta(\theta, Q) = \arg\max_{q} U(q, \theta) + W(q, \theta, Q).$$
(48)

Suppose the entrant offered the constant pricing scheme $p(q) = \varepsilon$, where ε is small. Under this pricing scheme, each customer would choose their intrinsic-value maximizing level of consumption $\alpha(\theta)$, and would realize surplus of $(U(\alpha(\theta), \theta) - \varepsilon)$. If this type of customer expected surplus of less than $(U(\alpha(\theta), \theta) - \varepsilon)$ from the monopolist's product, they would buy the entrant's product, and the entrant would receive non-zero profits. Therefore, in order to deter entry, the monopolist's pricing scheme must provide customers of type θ with a surplus of at least $(U(\alpha(\theta), \theta) - \varepsilon)$, for all $\varepsilon > 0$. Clearly, this cannot be achieved unless the pricing scheme provides customers of type θ with surplus of at least $U(\alpha(\theta), \theta)$. Since $U(\alpha(\theta), \theta)$ is the maximum surplus that a customer of type θ can get from the entrant's product under *any* pricing scheme, ensuring that customers get this level of surplus is both necessary and sufficient for the monopolist to deter entry.

As a consequence, when the fixed cost of entry is zero, deterring entry simply imposes a lower bound on the surplus each customer type must receive. Analytically, this is identical to the problem of choosing a pricing scheme with *type-dependent individual rationality* constraints. Therefore, setting $\hat{U}(\theta) = U(\alpha(\theta), \theta)$ in equation (7) ensures that any *Q*-feasible contract deters entry, and the definitions of all the other contracts in section 2.4 remain the same.

The monopolist's problem is to choose the optimal fulfilled-expectations contract, with $\hat{U}(\theta) = U(\alpha(\theta), \theta)$. In the following subsections, this problem is solved for *Q*-dependent and *q*, *Q*-dependent network effects. Rather than contrasting them with the base case, the contracts are compared to those derived in the absence of an entry threat (sections 4.3 and 4.4).

5.2 Entry deterrence with Q-proportional network effects

This subsection specifies the optimal fulfilled-expectations contracts when network value function depends on just gross consumption, and entry is successfully deterred. The following proposition establishes that the unique solution to the monopolist's problem in this case is to specify a quantity-independent (fixed-fee) pricing schedule:

Proposition 5 If $W(q, \theta, Q) = w(Q)$, then the optimal fulfilled-expectations contract that deters entry takes the form:

$$q^*(\theta) = \alpha(\theta); \tag{49}$$

$$\tau^*(\theta) = w(Q^*), \tag{50}$$

where $Q^* = \int_{\underline{\theta}}^{\overline{\theta}} \alpha(\theta) f(\theta) d\theta$. A contract of this form exists and is unique for any network value function w(Q).

Proposition 5 establishes that when deterring entry with network effects that depend on just gross consumption, the monopolist's optimal pricing scheme results in all customers choosing the level of consumption that maximizes total surplus⁷. Intuitively, the result is straightforward. A contract that separates any subset of types would need to induce consumption levels that are strictly lower than $\alpha(\theta)$ for all but the highest type in this subset. Ignoring the effect of reduced gross consumption on network value, this would result in a strict decrease in profits for the monopolist, since they would have to share some portion of the network value $w(Q^*)$ with the customers in this subset in order to ensure that their surplus is at least $U(\alpha(\theta), \theta)$. The effect of the accompanying reduction in gross consumption on network value accentuates this reduction in monopoly profits

⁷Note that since $W_1(q, \theta, Q) = 0$ in this case, $\beta(\theta, Q) = \alpha(\theta)$.

further. As a consequence, it is strictly profit-reducing to price-discriminate, and the monopolist offers the fixed-fee that maximizes profits. Since all customer get the same network value $w(Q^*)$, a fixed-fee contract that induces participation by any type induces participation by all types.

Relative to the model with no entry threat in Section 4.3, entry deterrence results in a strict increase in consumption, since $q^0(\theta) < \alpha(\theta)$ for all $\theta < \overline{\theta}$. Both the intrinsic value and the network value realized from each customer's consumption increase, and consequently, total surplus also increases. The surplus of all customer types increases, but the monopolist's profits are strictly lower.

5.3 Entry deterrence with q, Q-proportional network effects

This subsection analyzes optimal fulfilled-expectations contracts when network value function depends on both gross consumption and individual consumption, and entry is successfully deterred.

Let $q^m(\theta, Q)$ denote the Q-optimal contract in the absence of an entry threat (as in section 4.4). From Lemma 2, this allocation is defined for each θ by the necessary conditions

$$\frac{U_1(q^m(\theta, Q), \theta) + w(Q)}{U_{12}(q^m(\theta, Q), \theta)} = \frac{1 - F(\theta)}{f(\theta)},$$
(51)

and is unique for a fixed value of Q. Also, from Proposition 3, we know that there is an optimal fulfilled-expectations equilibrium – that is, there is a value of gross consumption such that

$$Q^{m} = \int_{\underline{\theta}}^{\overline{\theta}} q^{m}(\theta, Q^{m}) f(\theta) d\theta.$$
(52)

The following proposition establishes that the monopolist's pricing scheme results in individual consumption that is either of the form $q^m(\theta, Q)$, or that maximizes intrinsic value for the customer:

Proposition 6 Suppose $W(q, \theta, Q) = qw(Q)$. Assume that the uniqueness condition $w_1(Q) \leq -U_{11}(q,\overline{\theta})$ from proposition 3 is met. Define:

$$Q^{\alpha} = Q : q^{m}(\underline{\theta}, Q) = \alpha(\underline{\theta}), and$$
(53)

$$\hat{\theta}(Q) = \theta : q^m(\theta, Q) = \alpha(\theta).$$
(54)

(a) If $Q^{\alpha} \leq Q^{m}$, then the unique optimal fulfilled contract satisfies:

$$q^*(\theta) = q^m(\theta, Q^m) \tag{55}$$

$$\tau^*(\theta) = U(q^m(\theta), \theta) + q^m(\theta)w(Q^m) - U(\alpha(\theta), \theta) - \left[\int_{\underline{\theta}}^{\theta} (U_2(q^*(x), x) - U_2(\alpha(x), x))dx\right]$$
(56)

(b) If $Q^{\alpha} > Q^{m}$, then the unique optimal fulfilled contract satisfies:

$$q^*(\theta) = \alpha(\theta) \tag{57}$$

$$\tau^*(\theta) = \alpha(\theta)w(Q^*) \tag{58}$$

for $\theta \leq \hat{\theta}(Q^*)$, and

$$q^*(\theta) = q^m(\theta, Q^*) \tag{59}$$

$$\tau^{*}(\theta) = U(q^{*}(\theta), \theta) + q^{*}(\theta)w(Q^{*}) - U(\alpha(\theta), \theta) - \left[\int_{\hat{\theta}(Q)}^{\theta} (U_{2}(q^{*}(x), x) - U_{2}(\alpha(x), x))dx\right] (60)$$

for $\theta \geq \hat{\theta}(Q^*)$, where Q^* is the unique solution to:

$$Q = \int_{\underline{\theta}}^{\hat{\theta}(Q)} \alpha(\theta) f(\theta) d\theta + \int_{\hat{\theta}(Q)}^{\overline{\theta}} q^m(\theta, Q) f(\theta) d\theta.$$
(61)

Proposition 6 establishes that the same conditions that ensure uniqueness of the optimal fulfilledexpectations contract in the absence of an entry threat are sufficient to ensure uniqueness under the threat of entry. It also establishes that the optimal fulfilled-expectations contract that deters entry is a subset of two contracts – the one specified in Proposition 2, and the feasible fulfilledexpectations contract which implements allocations of $a(\theta)$.

For the lowest type, if $q^m(\underline{\theta}, Q^m) > \alpha(\underline{\theta})$, an immediate corollary of the proposition is that the presence of the entry threat does not change the individual consumption of any of the types (since $q^m(\underline{\theta}, Q^m) > \alpha(\underline{\theta})$ implies that $Q^\alpha < Q^m$). This is likely to happen when the marginal network value w(Q) is high relative to intrinsic value, or equivalently, if network effects are substantial for all types.

Under the conditions of part (b) of the proposition, the threat of entry induces an increase in individual consumption. However, even in this case, $\hat{\theta}(Q^*)$ is always an interior point of $[\underline{\theta}, \overline{\theta}]$. This implies that the increase in individual consumption will be for a subset of 'lower' types, and for any w(Q) > 0, there will always be a subset of higher types whose individual consumption remains unchanged. Therefore, while total surplus increases when entry is deterred, all of the increase comes from a lower subset of customers types.

All customers of type θ get surplus at least equal to $U(\alpha(\theta), \theta)$, which implies that they capture all of the intrinsic value that they create. Moreover, the customers whose consumption does not change – that is, the ones who continue to purchase $q^m(\theta, Q^m)$ – capture a fraction of the network value that they create. Since $U(\alpha(\theta), \theta) > U(q^*(\theta), \theta)$ for $q^*(\theta) > \alpha(\theta)$, the monopolist needs to give up network value to the customer if they raise consumption beyond $\alpha(\theta)$. The negative terms in square brackets at the end of equations (56) and (60) represent the surplus type θ gets beyond $U(\alpha(\theta), \theta)$, which implies that these customers are capturing a fraction over and above this minimum level, despite the fact that there is no heterogeneity in marginal network value w(Q)across customer types.

6 Discussion and Conclusion

A number of new results relating to the pricing of network goods have been derived in Sections 4 and 5. This section discusses some of these results, examines some of the model's assumptions, and concludes with an outline of open questions raised by the analysis.

6.1 Discussion of results

For a number of network goods, customer usage is in variable quantities, network value may be based on individual consumption, and may vary across customers. This paper provide a set of results which can form the basis for pricing policy for network goods of this kind. In addition, a substantial fraction of empirical work on network externalities (for instance, Gandal, 1995, Brynjolfsson and Kemerer, 1996, Forman, 2001) have focused on industries (databases, spreadsheets, LAN equipment) in which producers with monopoly power routinely offer nonlinear pricing schedules, sell variable quantities to customers, and potentially price to deter entry. Towards this end, the results of this paper would form a more robust theoretical basis for subsequent empirical work which aims to estimate the extent and implications of network effects in these industries

When network effects do not vary across customers, Proposition 1 indicates that the network effect causes no change in consumption, and that all surplus from the network effects is appropriated by the monopolist. When there is a threat of entry in the same setup, Proposition 5 shows that pricing changes substantially – a fixed fee is offered to all customer types – and that the threat of entry results in the outcome being socially optimal. While this is the simplest model of network value, it could apply to industries in which the primary network value stems from a fixed-cost reduction – for instance, the cost of finding the appropriate hosting infrastructure, or qualified technical support.

The pricing scheme in the latter case is similar to the cost-plus-fixed-fee results obtained in a duopoly model by Rochet and Stole (2001). In addition, the outcome suggests that if competing products are anything but perfectly compatible, any oligopoly outcome will always be socially inferior to the monopoly outcome, so long as the entry threat is maintained. In other words, from a policy perspective, ensuring a credible threat of entry is more socially efficient than actually

inducing entry.

When the value from the network effects varies with individual consumption, Proposition 2 indicates that the network effects induce an increase in consumption across all customer types, so long as the marginal network value is equal across types. In any model of price screening, there is always a trade-off between value creation and price discrimination, and the consumption of lower customer types is limited by the monopolist' desire to capture as much surplus as possible. The value creation aspect is accentuated further when there are network effects, since increases in consumption from any customer type increases the surplus created by all customer types. However, the trade-off still exists, and while pricing is redesigned to induce consumption increases from both lower and higher customer types, the consumption of lower types is still at an inefficient level. Moreover, sufficient heterogeneity in the marginal network value across types actually reduces the consumption of lower types, as indicated by Propositions 3 and 4, even when the marginal network value is positive for these types – this effect is driven entirely by price discrimination. Consequently, rather than benefiting all participants, certain kinds of network effects can actually harm lower-usage customers, and result in the distribution of customer surplus being skewed further towards high-end customers.

When network value depends on individual consumption as well as gross consumption, the effects of an entry threat are less pronounced that those predicted by Proposition 5. In fact, as shown in Proposition 6, the threat of entry may have no effect on consumption or surplus, and may merely results in a price change that redistributes surplus between the monopolist and its customers. Note that this occurs even when entry is not blockaded. This outcome is most likely when, relative to marginal intrinsic value, marginal network value is fairly high across all customers. When the entry threat does induce a change in total surplus, it stems from consumption disparities across different types of customers that stem from network effects can be partially mitigated by policy that creates a credible entry threat. In addition to reducing the differences in consumption across customers, there will be accompanying increases in surplus for all customers. While the outcome never maximizes total surplus, it is still likely that it is more efficient than an oligopoly with incompatible products.

The analysis of the division of surplus created by the network effects suggests that in the absence of an entry threat, the surplus captured by customers comes entirely from the network value created by changes in consumption. While the paper does not explore or model their source, the network effects displayed by the products used as examples to motivate the analysis are predominantly what would be called indirect network effects (or pecuniary externalities), and consequently, a discussion of the fact that the network 'externalities' are not internalized by the customers who create them is unlikely to be useful.

6.2 Discussion of assumptions

The sequence of events specified in section 2.3 assumes that all customers have identical expectations of gross consumption. Since all customers are rational, this is not restrictive – everyone has access to all the information needed to compute the expected consumption, and once the monopolist has specified prices, there is no residual uncertainty about what each type will demand. Clearly, in equilibrium, all customers must have the same expectation (the correct one).

However, compared to the standard model of nonlinear pricing, the model in this paper places a higher computational burden on the customers. Each customer has to know $F(\theta)$, compute the optimal consumption not just for themselves but for all customer types, and them calculate the gross consumption. It may be likely that customers of network goods cannot actually compute the true gross consumption immediately, either due to a lack of information, or due to bounds on processing capabilities. This suggests that there may be a multi-period adjustment process, in which customers iteratively make a series of guesses, which converge to the fulfilled-expectations equilibrium outcome. Alternately, customers may learn the distribution of types from the pricing schedule over time. Formalizing these notions remains (early-stage) work in progress.

The assumption that $U(q, \theta)$ has a finite maximum q for all θ is non-standard. However, given that marginal costs are zero in the model, it is necessary in order to get a bounded solution. It is also a reflection of reality – that customers do stop using zero marginal price products at a finite level, typically due to the presence of resource constraints, and substitute uses for shared resources, as discussed in Section 2.1.

In addition, slightly modified versions of all of the results in this paper continue to hold under the standard assumption of unbounded utility and convex costs. Consider, for instance, a (standard) specification in which customer utility is $\tilde{U}(q,\theta)+W(q,\theta,Q)$, $\tilde{U}_1(q,\theta) > 0$ for all q, and the provision of quantity q to each customer had a positive cost c(q), where $c_1(q) > 0$, $c_{11}(q) > 0$. If one defined the surplus function:

$$U(q,\theta) = U(q,\theta) - c(q),$$

then $U(q, \theta)$ would have the same properties as it does in this model. More importantly, all the expressions for $q^*(\theta)$ derived in the model would continue to be valid, and so would all the expressions for $\tau^*(\theta)$, if it is treated as the optimal markup rather than the optimal price. In other words, the optimal contracts would be $q^*(\theta), (\tau^*(\theta) + c(q^*(\theta)))$, with the same expressions for $q^*(\theta)$ and $\tau^*(\theta)$ derived in this paper.

To guarantee uniqueness, many of the propositions have specified conditions that need to be

imposed on $W(q, \theta, Q)$ – in general, they restrict the rate of change of marginal network value $W_1(q, \theta, Q)$ with increases in Q. When network effects are a primary source of value, these conditions may no longer be reasonable. However, all of the properties of the contracts derived in Propositions 1 through 5 do not depend on uniqueness. Moreover, if there are multiple optimal fulfilled-expectations equilibria, all the monopolist needs to do is choose the one with the highest profits – expectations are formed after prices are posted, and any price schedule has at most one optimal fulfilled-expectations equilibrium. Proposition 6 does rely on uniqueness, but a slightly modified version continues to hold if one assumes that the monopolist always chooses the highest-profit contract.

6.3 Concluding remarks

Industries in which products display network effects are often natural monopolies, especially when competing products are likely to be incompatible. Moreover, in at least one case (desktop operating systems), entry-deterrence appears to play a significant role in pricing. As a consequence, the analysis of an entry-deterring monopolist is likely to be of substantial importance for these industries. In light of the results obtained in this paper, one natural question that arises is how the presence of a non-zero entry cost affects outcomes. While it is clear that monopoly profits will increase, and entry deterrence will still be an optimal strategy, it is likely that the monopolist's profits will increase by less than the entry cost. The exact nature of how this affects consumption and welfare remains an open question.

Another set of open questions relate to the effect of relaxing the assumption that when network effects are type-dependent, the intrinsic value of the product and the value from the network effects are perfectly correlated. Since customer utility is separable in these two sources of value, one could model the network good as a multiproduct bundle, and characterize customers using a two-dimensional type vector, drawing on Armstrong (1996) and Rochet and Chone (1998). A particularly interesting case might be when the components of the vector are negatively correlated – for instance, in an electronic marketplace, where the intrinsic value depends on the frequency of trade while the value from network effects depends on the dollar value of the goods being traded. This would introduce countervailing incentives, and may lead to quantity-independent fixed fees being optimal.

Finally, the model of pricing under the threat of entry suggests the feasibility of analyzing a general model of nonlinear pricing for competing network goods. In the absence of network effects, and with zero marginal costs, results from Mandy (1992) indicate that in a duopoly, the only equilibrium pair of contracts involves both firms choosing a price of zero for all quantities. However, if customers expect the competing products to have different levels of gross consumption in equilibrium, they would view them as vertically differentiated products, which may facilitate non-zero equilibrium prices, as in Stole (1995). Similar issues have been analyzed in a model of coalition formation by Economides and Flyer (1998). The results from Section 4 suggest that in a general model, the equilibrium profits of the smaller network are unlikely to be non-zero. It is also possible that allowing mixed-strategies may lead to more interesting equilibrium outcomes, and recent results from Rochet and Stole (2001) facilitate an analysis of this kind. I hope to address some of these questions in the near future.

References

- 1. Armstrong, M. "Multidimensional Price Screening." Econometrica Vol. 64 (1996), pp. 51-75.
- Cabral, L., Salant, D., and Woroch, G., "Monopoly Pricing with Network Externalities." International Journal of Industrial Organization, Vol. 17 (1999), 199-214.
- Dewan, S. and Mendelson, H. "User Delay Costs and Internal Pricing for a Service Facility." Management Science Vol. 36 (1990), pp. 1502-1517.
- Dixit, A. K. "The Role of Investment in Entry Deterrence." *Economic Journal* Vol. 90 (1980), pp. 95-106.
- Economides, N. "The Economics of Networks." International Journal of Industrial Organization Vol. 14 (1996a), pp.673-699.
- Economides, N. "Network Externalities, Complementarities and Invitations to Enter." European Journal of Political Economy Vol. 12 (1996b), pp. 211-233.
- Economides, N., and Flyer, F. "Technical Standards Coalitions for Network Goods." Annales d'Economie et de Statistique Vol. 49/50 (1998), pp. 361-380.
- Farrell, J. and Saloner, G. "Standardization, Compatibility, and Innovation." Rand Journal of Economics Vol. 16, (1985), pp. 70-83.
- 9. Forman, C. "The Effects of Compatibility on Buyer Behavior in the Market for Computer Networking Equipment." Working Paper, Carnegie-Mellon University, 2001.
- 10. Fudenberg, D. and Tirole, J. Game Theory. Cambridge: MIT Press, 1991.
- Fudenberg, D. and Tirole, J. "Pricing a Network Good to Deter Entry." Journal of Industrial Economics, Vol. 48 (2000), pp. 373-390.

- Hall, C. E. and Hall, R. E. "Towards a More Precise Quantification of the Effects of Microsoft's Conduct', *American Economic Review* Vol. 92 (2002), pp. 188-191.
- Jullien, B. "Participation Constraints in Adverse Selection Problems." Journal of Economic Theory Vol. 93 (2000), 1-47.
- Jullien, B. "Competing in Network Industries: Divide and Conquer." Working Paper, IDEI, University of Toulouse, 2001.
- Katz, M. and Shapiro, C. "Network Externalities, Competition and Contracting." American Economic Review Vol. 75 (1985), pp. 424-440.
- Lewis, T. and Sappington, D. "Countervailing Incentives in Agency Problems." Journal of Economic Theory Vol. 82 (1989), pp. 294-313
- Mandy, D. "Nonuniform Bertrand Competition." *Econometrica* Vol. 60 (1992), pp. 1293-1330.
- Maskin, E. and Riley, J. "Monopoly with incomplete information." Rand Journal of Economics Vol. 15 (1984), pp. 171-196.
- Mendelson, H. "Pricing Computer Services: Queuing Effects." Communications of the ACM Vol. 28 (1985), pp. 312-321.
- Mendelson, H. and Whang, S. "Optimal Incentive-Compatible Priority Pricing for the M/M/1 queue." Operations Research Vol. 38 (1990), pp. 870-883.
- Radner, R. "Equilibrium Under Uncertainty." Chapter 20 in K. Arrow and M. Intriligator, eds., *Handbook of Mathematical Economics Vol. 2.* Amsterdam: North Holland Publishing Company, 1982.
- Rochet, J., and Chone, P. "Ironing, Sweeping and Multidimensional Screening." *Economet*rica Vol. 66 (1998), pp. 783-826
- 23. Rochet, J., and Stole, L. "Nonlinear Pricing with Random Participation." forthcoming, *Review of Economic Studies* (2001).
- Rohlfs, J. "A Theory of Interdependent Demand for a Communication Service." Bell Journal of Economics Vol. 10 (1974), pp. 16-37.
- 25. Salanié, B. The Economics of Contracts: A Primer. Cambridge: MIT Press, 1997.
- Schmalensee, R. Written testimony, Civil Action No. 98-1232 United States versus Microsoft, 1999. Available at http://www.microsoft.com/presspass/trial/schmal.asp.

- Segal, I., and Whinston, M. "Robust Predictions for Bilateral Contracting with Externalities." Working Paper #0027, Center for the Study of Industrial Organization, Northwestern University, 2001.
- Sibley, D., and Sriganesh, P. "Multiproduct Nonlinear Pricing with Multiple Taste Characteristics." *Rand Journal of Economics* Vol. 28 (1997), pp. 684-707.
- Spence, A. M. "Entry, Capacity, Investment and Oligopolistic Pricing." Bell Journal of Economics Vol. 8 (1977), pp. 534-547.
- Stole, L. "Nonlinear Pricing and Oligopoly." Journal of Economics and Management Strategy Vol. 4 (1995), pp. 529-562.
- Sundararajan, A. "Nonlinear Pricing of Information Goods." Working Paper #IS-02-01, Center for Digital Economy Research, New York University, 2002.
- Varian, H. and Shapiro, C. Information Rules. Cambridge: Harvard Business School Press, 1998.
- Westland, C. "Congestion and Network Externalities in the Short-Run Pricing of Information Systems Services." *Management Science* Vol. 38 (1992), pp. 992-1009.
- 34. Wilson, R. Nonlinear Pricing. New York: Oxford University Press, 1993.

A Appendix: Proofs

Proof of Lemma 2

(Note: This result can be directly inferred from the fact that for a fixed Q, the function $U(q, \theta) + W(q, \theta, Q)$ satisfies all the conditions in q and θ that a utility function needs to satisfy to ensure a unique, fully separating solution. For completeness, the details of the proof are provided below. Readers acquainted with the standard adverse selection theory with continuous types will find the arguments and sequence of analysis familiar, and can skip the details).

Given a expectation of gross consumption Q, any Q-feasible contract $q^F(t, Q), \tau^F(t, Q)$ satisfies [IC] is satisfied if:

$$\theta = \arg\max_{t \in [\underline{\theta}, \overline{\theta}]} U(q^F(t, Q), \theta) + W(q^F(t, Q), \theta, Q) - \tau^F(t, Q),$$
(62)

for all θ . The necessary and sufficient conditions for (62) are:

$$[U_1(q^F(t,Q),\theta) + W_1(q^F(t,Q),\theta,Q)]q_1^F(t,Q) - \tau_1^F(t,Q) = 0 \ \forall \theta;$$
(63)

$$[U_{11}(q^F(t,Q),\theta) + W_{11}(q^F(t,Q),\theta,Q)](q_1^F(t,Q))^2 + [U_1(q^F(t,Q),\theta) + W_1(q^F(t,Q),\theta,Q)]q_{11}^F(t,Q) - \tau_{11}^F(t,Q) \le 0 \ \forall \theta.$$
(64)

Differentiating (63) with respect to θ and substituting (64) yields the modified sufficient condition:

$$[U_{12}(q^F(t,Q),\theta) + W_{12}(q^F(t,Q),\theta,Q)]q_1^F(t,Q) \ge 0 \ \forall \theta.$$
(65)

By assumption, both $U_{12}(q,\theta)$ and $W_{12}(q,\theta,Q)$ are strictly positive, which means that (63) and (65) reduce to:

$$\tau_1^F(t,Q) = [U_1(q^F(t,Q),\theta) + W_1(q^F(t,Q),\theta,Q)]q_1^F(t,Q),$$
(66)

$$q_1^F(t,Q) \ge 0, \tag{67}$$

for all θ . Under $q^F(t, Q), \tau^F(t, Q)$, the surplus of type θ is

$$s(\theta) = U(q^F(\theta, Q), \theta) + W(q^F(\theta, Q), \theta, Q) - \tau^F(\theta, Q).$$
(68)

Differentiating (68) with respect to θ , and substituting (66) yields:

$$s_1(\theta) = U_2(q^F(\theta, Q), \theta) + W_2(q^F(\theta, Q), \theta, Q).$$
(69)

Since reservation utility $\hat{U}(\theta) = 0$ for all types, if IR is satisfied for the lowest type $\underline{\theta}$, it is satisfied for all others. Therefore, $s(\underline{\theta}) = 0$, and

$$s(\theta) = \int_{x=\underline{\theta}}^{\theta} [U_2(q^F(x,Q),x) + W_2(q^F(x,Q),x,Q)]dx.$$

$$(70)$$

From (68), we know that

$$\tau^{F}(\theta, Q) = U(q^{F}(\theta, Q), \theta) + W(q^{F}(\theta, Q), \theta, Q) - s(\theta),$$
(71)

which when combined with (70) means that the objective function whose maximizer is the optimal contract $q(\theta, Q), \tau(\theta, Q)$ can be written as:

$$\int_{\theta=\underline{\theta}}^{\overline{\theta}} \{U(q^{F}(\theta,Q),\theta) + W(q^{F}(\theta,Q),\theta,Q) \\
-\left[\int_{x=\underline{\theta}}^{\theta} U_{2}(q^{F}(x,Q),x) + W_{2}(q^{F}(x,Q),x,Q)dx]\}f(\theta)d\theta.$$
(72)

Integrating the second part of (72) by parts and rearranging yields:

$$q(\theta, Q) = \arg \max_{q^F(\theta, Q)_{\theta} = \underline{\theta}} \int_{-[U_2(q^F(\theta, Q), \theta) + W_2(q^F(\theta, Q), \theta, Q)]} \int_{-[U_2(q^F(\theta, Q), \theta) + W_2(q^F(\theta, Q), \theta, Q)]} \int_{-[H_2(\theta)]} \frac{1}{f(\theta)} f(\theta) d\theta,$$
(73)

subject to $q(\theta, Q) \ge 0$, and that

$$\tau(\theta, Q) = U(q(\theta, Q), \theta) + W(q(\theta, Q), \theta, Q) - \int_{x=\underline{\theta}}^{\theta} [U_2(q(x, Q), x) + W_2(q(x, Q), x, Q)] dx.$$
(74)

If the unconstrained problem has a unique solution for which $q(\theta, Q) \ge 0$, then this is the solution to the constrained problem as well.

Define

$$H(\theta) = \frac{1 - F(\theta)}{f(\theta)} \tag{75}$$

First-order conditions for the unconstrained problem are therefore:

$$U_1(q(\theta, Q), \theta) + W_1(q(\theta, Q), \theta, Q) = [U_{12}(q(\theta, Q), \theta) + W_{12}(q(\theta, Q), \theta, Q)]H(\theta) \ \forall \theta,$$
(76)

and are sufficient if the point-wise profit function:

$$\pi(q,\theta,Q) = U(q,\theta) + W(q,\theta,Q) - [U_2(q,\theta) + W_2(q,\theta,Q)]H(\theta)$$
(77)

is strictly concave in q. Differentiating (77) with respect to q twice yields:

$$\pi_{11}(q,\theta,Q) = U_{11}(q,\theta) + W_{11}(q,\theta,Q) - [U_{112}(q,\theta) + W_{112}(q,\theta,Q)]H(\theta),$$
(78)

which verifies that $\pi(q, \theta, Q)$ is strictly concave, since $U_{11} < 0$, $W_{11} \le 0$, $U_{112} \ge 0$ and $W_{112} \ge 0$. This ensures that for the unconstrained problem, first-order conditions (76) yield the unique solution. These conditions can be rearranged as:

$$\frac{U_1(q(\theta, Q), \theta) + W_1(q(\theta, Q), \theta, Q)}{[U_{12}(q(\theta, Q), \theta) + W_{12}(q(\theta, Q), \theta, Q)]} = \frac{1 - F(\theta)}{f(\theta)}.$$
(79)

Now, differentiating both sides of (76) with respect to θ yields:

$$\begin{aligned} & [U_{11}(q(\theta,Q),\theta) + W_{11}(q(\theta,Q),\theta,Q)]q_1(\theta,Q) + [U_{12}(q(\theta,Q),\theta) + W_{12}(q(\theta,Q),\theta,Q)] \ (80) \\ &= [U_{112}(q(\theta,Q),\theta) + W_{112}(q(\theta,Q),\theta,Q)]H(\theta)q_1(\theta,Q) \\ &+ [U_{122}(q(\theta,Q),\theta) + W_{122}(q(\theta,Q),\theta,Q)]H(\theta) \\ &+ [U_{12}(q(\theta,Q),\theta) + W_{12}(q(\theta,Q),\theta,Q)]H(\theta) \end{aligned}$$

which implies that:

$$q_{1}(\theta,Q) = \frac{[U_{12}(q(\theta,Q),\theta) + W_{12}(q(\theta,Q),\theta,Q)][1 - H_{1}(\theta)] - [U_{122}(q(\theta,Q),\theta) + W_{122}(q(\theta,Q),\theta,Q)]H(\theta)}{[U_{112}(q(\theta,Q),\theta) + W_{112}(q(\theta,Q),\theta,Q)]H(\theta) - [U_{11}(q(\theta,Q),\theta) + W_{11}(q(\theta,Q),\theta,Q)]}$$
(81)

Since $\pi(q, \theta, Q)$ has been shown to be strictly concave in q, the denominator of (81) is strictly positive. Also, the reciprocal of the hazard rate is non-increasing in θ , which implies that $H_1(\theta) \leq 0$.

Therefore, so long as both $U_{122}(q,\theta)$ and $W_{122}(q,\theta)$ are both non-positive, the numerator of (81) is strictly positive. Consequently, it has been established that

$$q_1(\theta, Q) > 0, \tag{82}$$

which implies that the unique unconstrained solution $q(\theta, Q)$ always satisfies the constraint, and completes the proof.

Proof of Lemma 3

Suppose $q^*(\theta), \tau^*(\theta)$ is an optimal fulfilled-expectations contract. From the definition of an optimal fulfilled-expectations contract, the contract defined by:

$$Q^* = \int_{\underline{\theta}}^{\overline{\theta}} q(\theta) f(\theta) d\theta$$
(83a)

$$q(\theta, Q^*) = q^*(\theta) \tag{83b}$$

$$\tau(\theta, Q^*) = \tau^*(\theta) \tag{83c}$$

must be a Q-optimal contract. From Lemma 2, $q^*(\theta), \tau^*(\theta)$ must therefore satisfy:

$$\frac{U_1(q^*(\theta), \theta) + W_1(q^*(\theta), \theta, Q^*)}{[U_{12}(q^*(\theta), \theta) + W_{12}(q^*(\theta), \theta, Q^*)]} = \frac{1 - F(\theta)}{f(\theta)}.$$
(84)

$$\tau^*(\theta) = U(q^*(\theta), \theta) + W(q^*(\theta), \theta, Q^*) - \int_{x=\underline{\theta}}^{\theta} [U_2(q^*(x), x) + W_2(q^*(x), x, Q^*)] dx,$$
(85)

which completes the proof.

Before proceeding further, define the gross consumption function:

$$\Gamma(Q) = \int_{\underline{\theta}}^{\theta} q(\theta, Q) f(\theta) d\theta, \qquad (86)$$

where $q(\theta, Q)$ is part of the unique Q-optimal contract associated with an expected gross consumption Q. Two immediate consequences of this definition are:

Lemma 4 (a) For any Q-optimal contract $q(\theta, Q), \tau(\theta, Q)$, if $\Gamma(Q) = Q$, then the contract defined by

$$q^*(\theta) = q(\theta, Q); \tag{87}$$

$$\tau^*(\theta) = \tau(\theta, Q) \tag{88}$$

is an optimal fulfilled-expectations contract.

(b) $\Gamma(0) > 0$.

Part (a) follows directly from the definition of $\Gamma(Q)$ and of an optimal fulfilled-expectations contract, and part (b) from the fact that intrinsic value is always positive. We also need another result – the strict monotonicity of $\frac{U_1(q,\theta)}{U_{12}(q,\theta)}$, which follows from decreasing absolute risk aversion:

Lemma 5 If $\frac{d}{d\theta}\left(\frac{-U_{11}(q,\theta)}{U_1(q,\theta)}\right) < 0$, then $\frac{d}{dq}\left(\frac{U_1(q,\theta)}{U_{12}(q,\theta)}\right) < 0$.

Proof:

$$\frac{d}{d\theta}\left(\frac{-U_{11}(q,\theta)}{U_1(q,\theta)}\right) = \frac{-U_{112}(q,\theta)U_1(q,\theta) + U_{11}(q,\theta)U_{12}(q,\theta)}{(U_1(q,\theta))^2},\tag{89}$$

and

$$\frac{d}{dq}\left(\frac{U_1(q,\theta)}{U_{12}(q,\theta)}\right) = \frac{U_{11}(q,\theta)U_{12}(q,\theta) - U_{112}(q,\theta)U_1(q,\theta)}{(U_{12}(q,\theta))^2}.$$
(90)

The denominators of the RHS (89) and (90) are both strictly positive, and the numerators are identical. The result follows.

Proof of Proposition 1

Since $W(q, \theta, Q) = w(Q)$, it follows that $W_1(q, \theta, Q) = W_2(q, \theta, Q) = W_{12}(q, \theta, Q) = 0$. Therefore, from Lemma 3, the optimal fulfilled-expectations contract satisfies:

$$\frac{U_1(q^*(\theta),\theta)}{U_{12}(q^*(\theta),\theta)} = \frac{1 - F(\theta)}{f(\theta)},\tag{91}$$

and

$$\tau^*(\theta) = U(q^*(\theta), \theta) + w(Q^*) - \int_{x=\underline{\theta}}^{\theta} [U_2(q^*(x), x)] dx, \qquad (92)$$

where $Q^* = \int_{\underline{\theta}}^{\overline{\theta}} q^*(\theta) f(\theta) d\theta$. Comparing (91) with (30) yields

$$q^*(\theta) = q^0(\theta), \tag{93}$$

since Lemma 5 has established that $\frac{U_1(q,\theta)}{U_{12}(q,\theta)}$ is strictly monotonic in q. Therefore, $Q^* = Q^0$. Consequently, comparing (92) with (31) yields:

$$\tau^*(\theta) = \tau^0(\theta) + w(Q^0). \tag{94}$$

Similarly, from Lemma 2, for any expectation of gross consumption Q, the individual consumption in any Q-optimal contract satisfies:

$$\frac{U_1(q(\theta, Q), \theta)}{U_{12}(q(\theta, Q), \theta)} = \frac{1 - F(\theta)}{f(\theta)},\tag{95}$$

which implies that

$$q(\theta, Q) = q^0(\theta) \tag{96}$$

for all Q and θ . Therefore, (86), (91) and (96) imply that $\Gamma(Q) = Q^0$ for all Q. Clearly, $\Gamma(Q)$ always has a unique fixed point Q^0 , which completes the proof.

Proof of Proposition 2

(a) Since $W(q, \theta, Q) = qw(Q)$, $W_1(q, \theta, Q) = w(Q)$, and $W_2(q, \theta, Q) = W_{12}(q, \theta, Q) = 0$. The necessary conditions now follow directly from Lemma 3.

(b) From Lemma 2, for any Q, individual consumption in the unique Q-optimal contract satisfies:

$$\frac{U_1(q(\theta,Q),\theta) + w(Q)}{U_{12}(q(\theta,Q),\theta)} = \frac{1 - F(\theta)}{f(\theta)}$$

$$\tag{97}$$

Existence: If w(Q) is bounded, this implies that $U_1(q(\theta, Q), \theta)$ is bounded for all θ , which in turn implies that $q(\theta, Q)$ is bounded for all θ , since $U_{11}(q, \theta) < 0$. Therefore $\Gamma(Q)$ is bounded. Since $\Gamma(0) > 0$, this implies that a fixed point for $\Gamma(Q)$ exists.

Uniqueness: Differentiating both sides of (97) with respect to Q and rearranging yields:

$$q_2(\theta, Q) = \frac{w_1(Q)}{U_{112}(q(\theta, Q), \theta) \frac{1 - F(\theta)}{f(\theta)} - U_{11}(q(\theta, Q), \theta)}.$$
(98)

Since $U_{112}(q(\theta, Q), \theta) \ge 0$, this implies that

$$q_2(\theta, Q) \le \frac{w_1(Q)}{-U_{11}(q(\theta, Q), \theta)}.$$
(99)

From the conditions of the proposition for uniqueness, we know that $w_1(Q) < -U_{11}(q,\theta)$, which when combined with (99) implies that

$$q_2(\theta, Q) < 1 \tag{100}$$

Now, differentiating both sides of (86) with respect to Q yields

$$\Gamma_1(Q) = \int_{\underline{\theta}}^{\overline{\theta}} q_2(\theta, Q) f(\theta) d\theta, \qquad (101)$$

which when combined with (100), implies that $\Gamma_1(Q) < 1$ for all Q. This in turn implies that $\Gamma(Q)$ is a contraction, and since $\Gamma(0) > 0$, it has a unique and strictly positive fixed point.

(c) When w(Q) > 0, (39) implies that

$$\frac{U_1(q^*(\theta), \theta)}{U_{12}(q^*(\theta), \theta)} < \frac{1 - F(\theta)}{f(\theta)}$$
(102)

From Lemma 1, we know that

$$\frac{U_1(q^0(\theta), \theta)}{U_{12}(q^0(\theta), \theta)} = \frac{1 - F(\theta)}{f(\theta)}$$
(103)

and therefore

$$\frac{U_1(q^0(\theta),\theta)}{U_{12}(q^0(\theta),\theta)} > \frac{U_1(q^*(\theta),\theta)}{U_{12}(q^*(\theta),\theta)}$$
(104)

From Lemma 5, we know that $\frac{U_1(q,\theta)}{U_{12}(q,\theta)}$ is strictly decreasing in q for all θ , which when combined with (104) proves that $q^0(\theta) < q^*(\theta)$ for all θ .

(d) Consider the contract:

$$q^{FE}(\theta) = q^0(\theta) \tag{105a}$$

$$\tau^{FE}(\theta) = \tau^0(\theta) + q^0(\theta)W(Q^0)$$
(105b)

It is straightforward to establish that this is a *feasible* fulfilled-expectations contract. Under this contract, the monopolist's profits would be

$$\Pi = \int_{\underline{\theta}}^{\overline{\theta}} \tau^{0}(\theta) f(\theta) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} q^{0}(\theta) W(Q^{0}) f(\theta) d\theta$$
(106)

Using the definition of $s^n(\theta)$ from (16), this implies that

$$\Pi = \int_{\underline{\theta}}^{\overline{\theta}} \tau^{0}(\theta) f(\theta) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} s^{n}(\theta) f(\theta) d\theta$$
(107)

Profits under the optimal fulfilled expectations contract must be at least as high as Π . Based on (107), this yields:

$$\int_{\underline{\theta}}^{\overline{\theta}} \tau^*(\theta) f(\theta) d\theta \ge \int_{\underline{\theta}}^{\overline{\theta}} \tau^0(\theta) f(\theta) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} s^n(\theta) f(\theta) d\theta,$$

which proves the first part. Now denote the surplus of type θ under the optimal fulfilled-expectations contract as $s^*(\theta)$, and the surplus of type θ under the base case contract as $s^0(\theta)$. We know that

$$s^*(\theta) = \int_{x=\underline{\theta}}^{\theta} [U_2(q^*(x), x)] dx,$$

ė

and

$$s^{0}(\theta) = \int_{x=\underline{\theta}}^{\theta} [U_{2}(q^{0}(x), x)]dx.$$

Since $q^*(\theta) > q^0(\theta)$, and $U_{12}(q,\theta) > 0$, this implies that $s^*(\theta) > s^0(\theta)$ for all θ . Therefore, the monopolist does not appropriate all the surplus generated by the network effects, and the second part of the result follows.

Proof of Proposition 3

(a) Since $W(q, \theta, Q) = qw(Q, \theta)$, $W_1(q, \theta, Q) = w(Q, \theta)$, $W_2(q, \theta, Q) = qw_2(Q, \theta)$, and $W_{12}(q, \theta, Q) = w_2(Q, \theta)$. The necessary conditions now follow directly from Lemma 3.

b) From Lemma 2, for any Q, individual consumption in the unique Q-optimal contract satisfies:

$$\frac{U_1(q(\theta, Q), \theta) + w(Q, \theta)}{U_{12}(q(\theta, Q), \theta) + w_2(Q, \theta)} = \frac{1 - F(\theta)}{f(\theta)}$$
(108)

Existence: If $w(Q, \theta)$ is bounded, this implies that $U_1(q(\theta, Q), \theta)$ is bounded, which in turn implies that $q(\theta, Q)$ is bounded for all θ , since $U_{11}(q, \theta) < 0$. Therefore $\Gamma(Q)$ is bounded, and the result follows.

Uniqueness: Differentiating both sides of (108) with respect to Q and rearranging yields:

$$q_2(\theta, Q) = \frac{w_1(Q, \theta) - w_{12}(Q, \theta) \frac{1 - F(\theta)}{f(\theta)}}{U_{112}(q(\theta, Q), \theta) \frac{1 - F(\theta)}{f(\theta)} - U_{11}(q(\theta, Q), \theta)}.$$
(109)

Since $W_{123}(q, \theta, Q) \ge 0$, this implies that $w_{12}(Q, \theta) \ge 0$. Also, we know that $U_{112}(q(\theta, Q), \theta) \ge 0$. Therefore,

$$q_2(\theta, Q) \le \frac{w_1(Q, \theta)}{-U_{11}(q(\theta, Q), \theta)}.$$
(110)

From the conditions of the proposition for uniqueness, we know that $w_1(Q,\theta) < -U_{11}(q(\theta,Q),\theta)$, which when combined with (110) implies that

$$q_2(\theta, Q) < 1 \tag{111}$$

Now, differentiating both sides of (86) with respect to Q yields

$$\Gamma_1(Q) = \int_{\underline{\theta}}^{\overline{\theta}} q_2(\theta, Q) f(\theta) d\theta, \qquad (112)$$

which when combined with (111), implies that $\Gamma_1(Q) < 1$ for all Q. This in turn implies that $\Gamma(Q)$ is a contraction, and since $\Gamma(0) > 0$, it has a unique fixed point.

(c) Assume first that

$$\frac{U_1(q^*(\theta), \theta)}{U_{12}(q^*(\theta), \theta)} > \frac{w(Q^*, \theta)}{w_2(Q^*, \theta)}.$$
(113)

This implies that

$$\frac{U_1(q^*(\theta), \theta) + w(Q^*, \theta)}{U_{12}(q^*(\theta), \theta) + w_2(Q^*, \theta)} < \frac{U_1(q^*(\theta), \theta)}{U_{12}(q^*(\theta), \theta)},$$
(114)

which when combined with (43) yields:

$$\frac{U_1(q^*(\theta),\theta)}{U_{12}(q^*(\theta),\theta)} > \frac{1 - F(\theta)}{f(\theta)}.$$
(115)

Also, we know that

$$\frac{U_1(q^0(\theta),\theta)}{U_{12}(q^0(\theta),\theta)} = \frac{1 - F(\theta)}{f(\theta)}.$$
(116)

(115) and (116) imply that

$$\frac{U_1(q^0(\theta),\theta)}{U_{12}(q^0(\theta),\theta)} < \frac{U_1(q^*(\theta),\theta)}{U_{12}(q^*(\theta),\theta)},\tag{117}$$

which implies that $q^0(\theta) > q^*(\theta)$, since we know from Lemma 5 that $\frac{d}{dq}(\frac{U_1(q,\theta)}{U_{12}(q,\theta)}) < 0$.

Similarly, if

$$\frac{U_1(q^*(\theta),\theta)}{U_{12}(q^*(\theta),\theta)} < \frac{w(Q^*,\theta)}{w_2(Q^*,\theta)},\tag{118}$$

this implies that

$$\frac{U_1(q^*(\theta), \theta) + w(Q^*, \theta)}{U_{12}(q^*(\theta), \theta) + w_2(Q^*, \theta)} > \frac{U_1(q^*(\theta), \theta)}{U_{12}(q^*(\theta), \theta)},$$
(119)

and the converse result follows from an analogous sequence of steps.

Proof of Proposition 4

Differentiating $\frac{w(Q^*,\theta)}{w_2(Q^*,\theta)}$ with respect to θ yields:

$$\frac{d}{d\theta} \left(\frac{w(Q^*,\theta)}{w_2(Q^*,\theta)}\right) = 1 - \frac{w(Q^*,\theta)w_{22}(Q^*,\theta)}{(w_2(Q^*,\theta))^2}.$$
(120)

If $w(Q^*, \theta)$ is concave in θ , this implies that

$$\frac{d}{d\theta}\left(\frac{w(Q^*,\theta)}{w_2(Q^*,\theta)}\right) > 1.$$
(121)

Also,

$$\frac{d}{d\theta} \left(\frac{U_1(q^*(\theta), \theta)}{U_{12}(q^*(\theta), \theta)} \right) = 1 + \frac{U_{11}(q^*(\theta), \theta)q_1^*(\theta)}{U_{12}(q^*(\theta), \theta)} - \frac{U_{12}(q^*(\theta), \theta)q_1^*(\theta) + U_{122}(q^*(\theta), \theta)]}{(U_{12}(q^*(\theta), \theta))^2}.$$
(122)

It is straightforward (though tedious) to show that when one combines (122) with Lemma 5, the fact that $q_1^*(\theta) > 0$, and either of the conditions on $U(q, \theta)$ in the proposition, one gets:

$$\frac{d}{d\theta} \left(\frac{U_1(q^*(\theta), \theta)}{U_{12}(q^*(\theta), \theta)} \right) < 1.$$
(123)

(a) Equations (121) and (123) together imply that if

$$\frac{U_1(q^*(\underline{\theta}),\underline{\theta})}{U_{12}(q^*(\underline{\theta}),\underline{\theta})} < \frac{w(Q^*,\underline{\theta})}{w_2(Q^*,\underline{\theta})},\tag{124}$$

then $\frac{U_1(q^*(\theta), \theta)}{U_{12}(q^*(\theta), \theta)} < \frac{w(Q^*, \theta)}{w_2(Q^*, \theta)}$ for all $\theta > \underline{\theta}$. When combined with Proposition 3(c), the result follows.

(b) From (43), we know that

$$\frac{U_1(q^*(\overline{\theta}),\overline{\theta}) + w(Q^*,\overline{\theta})}{U_{12}(q^*(\overline{\theta}),\overline{\theta}) + w_2(Q^*,\overline{\theta})} = 0,$$
(125)

which implies that $U_1(q^*(\overline{\theta}), \overline{\theta}) < 0$, which in turn implies that

$$\frac{U_1(q^*(\overline{\theta}),\overline{\theta})}{U_{12}(q^*(\overline{\theta}),\overline{\theta})} < \frac{w(Q^*,\overline{\theta})}{w_2(Q^*,\overline{\theta})},\tag{126}$$

since $U_{12}(q,\theta)$, $w(Q,\theta)$ and $w_2(Q,\theta)$ are all positive. Now, since $q^*(\theta)$ is continuous, we know that both $\frac{U_1(q^*(\theta),\theta)}{U_{12}(q^*(\theta),\theta)}$ and $\frac{w(Q^*,\theta)}{w_2(Q^*,\theta)}$ are continuous. Therefore, if

$$\frac{U_1(q^*(\underline{\theta}),\underline{\theta})}{U_{12}(q^*(\underline{\theta}),\underline{\theta})} > \frac{w(Q^*,\underline{\theta})}{w_2(Q^*,\underline{\theta})},\tag{127}$$

then (126) and (127) imply that there is some $\hat{\theta} \in (\underline{\theta}, \overline{\theta})$ such that

$$\frac{U_1(q^*(\theta), \theta)}{U_{12}(q^*(\hat{\theta}), \hat{\theta})} = \frac{w(Q^*, \theta)}{w_2(Q^*, \hat{\theta})}.$$
(128)

Moreover, given (128), (121) and (123) imply that for $\theta < \hat{\theta},$

$$\frac{U_1(q^*(\theta), \theta)}{U_{12}(q^*(\theta), \theta)} > \frac{w(Q^*, \theta)}{w_2(Q^*, \theta)},$$
(129)

and for $\theta > \hat{\theta}$,

$$\frac{U_1(q^*(\theta),\theta)}{U_{12}(q^*(\theta),\theta)} < \frac{w(Q^*,\theta)}{w_2(Q^*,\theta)}.$$
(130)

The result now follows from Proposition 3(c).

Proof of Proposition 5

Consider any expectation of gross consumption Q, and any Q-feasible contract $q^F(\theta), \tau^F(\theta)$. [IR] implies that:

$$U(q^{F}(\theta), \theta) + w(Q) - \tau^{F}(\theta) \ge U(\alpha(\theta), \theta).$$
(131)

Since $U(q^F(\theta), \theta) \leq U(\alpha(\theta), \theta)$, this implies that

$$\tau^F(\theta) \le w(Q) \tag{132}$$

for all θ . Consequently, a pricing scheme that provides the monopolist with a payment of w(Q) from each type has equal or higher profits than any Q-feasible contract. The only incentive-compatible individual consumption under any contract of this kind is when customers of type θ are allocated $\alpha(\theta)$. This means that

$$q(\theta, Q) = \alpha(\theta); \tag{133}$$

$$\tau(\theta, Q) = w(Q) \tag{134}$$

is Q-optimal. Defining

$$Q^{\alpha} = \int_{\underline{\theta}}^{\theta} \alpha(\theta) f(\theta) d\theta, \qquad (135)$$

this implies that $\Gamma(Q) = Q^{\alpha}$ for all Q, which in turn implies that there is a unique optimal fulfilledexpectations contract, and the result follows.

Proof of Proposition 6

This proof introduces some new notation. The notation follows Jullien (2000), since Proposition 3 of that paper is used to establish part of this result.

For any expectation of gross consumption Q, define:

$$l(\gamma, \theta, Q) = \arg\max_{q} U(q, \theta) + qw(Q) - U_2(q, \theta) \frac{\gamma - F(\theta)}{f(\theta)}.$$
(136)

It follows immediately that

$$l(1,\theta,Q) = q^m(\theta,Q), \tag{137}$$

and that $l(0, \theta, Q) \ge \beta(\theta, Q)$ for all θ , which implies that $l(0, \theta, Q) > \alpha(\theta)$ for all θ .

Next, the unique incentive-compatible contract that implements $\hat{U}(\theta)$ – that is, the unique contract under which the surplus of type θ is $\hat{U}(\theta)$ – can be shown to be:

$$\hat{q}(\theta) = \alpha(\theta); \tag{138}$$

$$\hat{\tau}(\theta) = \alpha(\theta)w(Q). \tag{139}$$

Therefore, the set $\Theta = \{\theta : l(1, \theta, Q) \le \hat{q}(\theta) \le l(0, \theta, Q)\}$ reduces to:

$$\Theta = \{\theta : q^m(\theta, Q) \le \alpha(\theta)\}.$$
(140)

Also, define

$$\hat{\gamma}(\theta, Q) = \gamma : \hat{q}(\theta) = \arg\max_{q} U(q, \theta) + qw(Q) - U_2(q, \theta) \frac{\gamma - F(\theta)}{f(\theta)}$$
(141)

Since $\hat{q}(\theta) = \alpha(\theta)$, and $U_1(\alpha(\theta), \theta) = 0$, first-order conditions for (141) yield:

$$\hat{\gamma}(\theta, Q) = F(\theta) + \frac{f(\theta)w(Q)}{U_{12}(\alpha(\theta), \theta)}$$
(142)

Finally, define:

$$H(\gamma, \theta) = \frac{\gamma - F(\theta)}{f(\theta)}$$

The following lemma is straightforward, and is stated without proof.

Lemma 6 If $\frac{1-F(\theta)}{f(\theta)}$ is non-decreasing for all θ , then $H_2(\gamma, \theta) \leq 0$ for all θ, γ such that $H(\gamma, \theta) \geq 0$.

The next lemma can now be proved:

Lemma 7 If $\frac{\partial}{\partial \theta} \left(\frac{U_{11}(\alpha(\theta), \theta)}{U_{12}(\alpha(\theta), \theta)} \right) \leq 0$, then $l_2(\hat{\gamma}(\theta, Q), \theta, Q) \leq \hat{q}_1(\theta)$

Proof: By the definition of $\hat{\gamma}(\theta, Q)$ in (141), and of $l(\gamma, \theta, Q)$ in (136), we know that

$$l(\hat{\gamma}(\theta, Q), \theta, Q) = \hat{q}(\theta) = \alpha(\theta).$$
(143)

From the first order conditions for (136), we know that

$$U_1(l(\gamma,\theta,Q),\theta) + w(Q) = U_{12}(l(\gamma,\theta,Q),\theta) \frac{\gamma - F(\theta)}{f(\theta)}.$$
(144)

Differentiating both sides of (144) with respect to θ and rearranging yields:

$$l_{2}(\gamma,\theta,Q) = \frac{U_{12}(l(\gamma,\theta,Q),\theta)(1-H_{2}(\gamma,\theta)) - U_{122}(l(\gamma,\theta,Q),\theta)H(\gamma,\theta)}{U_{112}(l(\gamma,\theta,Q),\theta)H(\gamma,\theta) - U_{11}(l(\gamma,\theta,Q),\theta)}.$$
 (145)

Substituting in $\hat{\gamma}(\theta, Q)$ and using (143) yields:

$$l_2(\hat{\gamma}(\theta, Q), \theta, Q) = \frac{U_{12}(\alpha(\theta), \theta)(1 - H_2(\hat{\gamma}(\theta, Q), \theta)) - U_{122}(\alpha(\theta), Q), \theta)H(\hat{\gamma}(\theta, Q), \theta)}{U_{112}(\alpha(\theta), \theta)H(\hat{\gamma}(\theta, Q), \theta) - U_{11}(\alpha(\theta), \theta)}.$$
 (146)

Since $\hat{q}(\theta) = \alpha(\theta)$, and $U_1(\alpha(\theta), \theta) = 0$ by definition, it follows that:

$$\hat{q}_1(\theta) = \frac{U_{12}(\alpha(\theta), \theta)}{-U_{11}(\alpha(\theta), \theta)}.$$
(147)

Comparing equations (146) and (147), and using the fact that $U_{11}(q,\theta)U_{12}(q,\theta)H_2(\gamma,\theta) \ge 0$, it follows that $l_2(\hat{\gamma}(\theta,Q),\theta,Q) \le \hat{q}_1(\theta)$ if:

$$U_{12}((\alpha(\theta),\theta))U_{112}(\alpha(\theta),\theta) - U_{11}(\alpha(\theta),\theta)U_{122}(\alpha(\theta),\theta) \le 0,$$
(148)

which is precisely the condition implied by $\frac{\partial}{\partial \theta} \left(\frac{U_{11}(q,\theta)}{U_{12}(q,\theta)} \right) \leq 0$. The result follows.

It was shown in Section 5.1 that the problem of finding an optimal contract which deters entry was equivalent to finding an optimal contract with type-dependent participation constraints. The following result therefore holds, based on Jullien (2000):

Lemma 8 The Q-optimal contract which deters entry satisfies:

(a) If $\theta \in \Theta$,

$$q(\theta, Q) = \alpha(\theta);$$
(149)

$$\tau(\theta) = a(\theta)w(Q).$$

(b) If $\theta \notin \Theta$, then:

$$q(\theta,Q) = q^{m}(\theta,Q);$$

$$\tau(\theta) = U(q^{m}(\theta,Q),\theta) + q^{m}(\theta,Q)w(Q) - U(\alpha(\theta),\theta) - \left[\int_{\hat{\theta}(Q)}^{\theta} (U_{2}(q^{m}(x,Q),x) - U_{2}(\alpha(x),x))dx\right],$$
(150)

where $\Theta = \{\theta : q^m(\theta, Q) \le \alpha(\theta)\}$, and $\hat{\theta}(Q) = \theta : q^m(\theta, Q) = \alpha(\theta)$.

In addition, if Θ is empty, then for all θ :

$$q(\theta, Q) = q^{m}(\theta, Q),$$

$$\tau(\theta) = U(q^{m}(\theta, Q), \theta) + q^{m}(\theta, Q)w(Q) - U(\alpha(\theta), \theta) - [\int_{\underline{\theta}}^{\theta} (U_{2}(q^{m}(x, Q), x) - U_{2}(\alpha(x), x))dx].$$
(151)

Proof: Lemma 7 ensures that the problem of finding a Q-optimal contract satisfies all the conditions for Proposition 3 of Jullien (2000) to apply. The expressions for $q(\theta, Q)$ follow immediately. The expressions for $\tau(\theta, Q)$ follow by imposing incentive-compatibility.

Now, recall the definitions of Q^m, Q^α from the proposition, and $\Gamma(Q)$ from the proof of proposition 2:

$$Q^m = Q: Q = \int_{\theta}^{\overline{\theta}} q^m(\theta, Q) f(\theta) d\theta$$
(152)

$$Q^{\alpha} = Q: q^{m}(\underline{\theta}, Q) = \alpha(\underline{\theta}), \qquad (153)$$

$$\Gamma(Q) = \int_{\underline{\theta}}^{\theta} q^{m}(\theta, Q) f(\theta) d\theta$$
(154)

Clearly, Q^m is a fixed point of $\Gamma(Q)$. Also define $\hat{\Gamma}(Q)$ as

$$\hat{\Gamma}(Q) = \int_{\underline{\theta}}^{\hat{\theta}(Q)} \alpha(\theta) f(\theta) d\theta + \int_{\hat{\theta}(Q)}^{\overline{\theta}} q^m(\theta, Q) f(\theta) d\theta, \text{ if } \hat{\theta}(Q) \text{ exists;}$$
(155)

$$\hat{\Gamma}(Q) = \Gamma(Q)$$
 otherwise. (156)

Now, from (98). we know that

$$q_2^m(\theta, Q) = \frac{w_1(Q)}{U_{112}(q^m(\theta, Q), \theta) \frac{1 - F(\theta)}{f(\theta)} - U_{11}(q^m(\theta, Q), \theta)}.$$
(157)

The RHS of (157) is strictly positive since $w_1(Q) > 0$, $U_{112}(q,\theta) \ge 0$, and $U_{11}(q,\theta) < 0$. Therefore, $q_2^m(\theta,Q) > 0$ for all Q. As a consequence, if $Q^{\alpha} < Q^m$, then $q^m(\underline{\theta},Q^{\alpha}) < q^m(\underline{\theta},Q^m)$, which in turn implies that $q^m(\underline{\theta},Q^m) > \alpha(\underline{\theta})$. The set Θ is therefore empty, and this establishes part (a), based on Lemma 8.

Now, differentiating both sides of (155) and (156) with respect to Q yields:

$$\hat{\Gamma}_1(Q) = \int_{\hat{\theta}(Q)}^{\overline{\theta}} q_1^m(\theta, Q) f(\theta) d\theta, \text{ if } \hat{\theta}(Q) \text{ exists;}$$
(158)

$$\hat{\Gamma}_1(Q) = \Gamma_1(Q)$$
 otherwise. (159)

Since $q_2^m(\theta, Q) > 0$, this implies that $\hat{\Gamma}_1(Q) \leq \Gamma_1(Q)$, and the inequality is strict if $\hat{\theta}(Q) > \underline{\theta}$. Consequently, under the conditions for uniqueness in Proposition 2, both $\Gamma(Q)$ and $\hat{\Gamma}(Q)$ have unique strictly positive fixed points.

If $Q \ge Q^{\alpha}$, $\hat{\Gamma}(Q) = \Gamma(Q)$. As a consequence, if $Q^{\alpha} > Q^{m}$, this means that the fixed point of $\hat{\Gamma}(Q)$ has to lie in (Q^{m}, Q^{α}) , because we know that $\hat{\Gamma}(Q) > \Gamma(Q)$ for $Q < Q^{\alpha}$, which means it cannot have a fixed point at Q^{m} , which in turn implies that if it has a fixed point greater than Q^{α} , this violates the uniqueness of the fixed point of $\Gamma(Q)$.

Using the fact that the unique optimal fulfilled-expectation equilibrium has gross consumption that is the fixed point of $\hat{\Gamma}(Q)$, part (b) of the result follows. This completes the proof.