The Economics of Number Portability: Switching Costs and Two-Part Tariffs *

Reiko Aoki† John Small ‡

August 2010 §

Abstract

This paper interprets number portability as a reduction of switching costs in a model of competition between telephone companies. We identify several cases by their cost and demand characteristics and show that social benefit of number portability are not guaranteed. Analysis using two-part tariff highlights the effect of how the technological cost of switching cost reduction effects the final market allocation.

*We gratefully acknowledge the comments and suggestions of Justus Haucap, John Panzar, Julian Wright and other participants in seminars at University of Auckland, Northwestern University, and Osaka University Institute of Social and Economic Research.

†Center for Intergenerational Studies, Institute of Economic Research, Hitotsubashi University, aokirei@ier.hit-u.ac.jp

‡Covec, Auckland, New Zealand, john.small@covec.co.nz

§Originally Aoki, R., Small, J. "The Economics of Number Portability: Switching Costs and Two Part Tariffs". CRNEC Working Paper, University of Auckland, 1999.

1 Introduction

The introduction of number portability is one of the most active policy challenges facing the telecommunications industry worldwide. The ability to retain a telephone number while switching carrier is known as *operator portability* and is the primary focus of this paper. Operator portability has become regarded as an almost essential pre-condition for local loop competition and regulators in several jurisdictions have already set timetables for its introduction. In the United States of America, for example, the Federal Communications Commission has mandated number portability, and laid down performance criteria for long term solutions which effectively require intelligent networks. Denmark and Hong Kong have already implemented operator portability, although by quite different methods. A very different approach has been used in New Zealand, where the policy has effectively been devolved to the telecommunications industry.[1]

In most countries policy towards number portability lies somewhere between these extremes, with analysts being interested in evaluating the merits of imposing a porting requirement. The overall aim of this paper is to assist this policy formation process by providing a welfare analysis of portability.

The great variety of methods by which portability can be delivered is indicative of the pace of technological change in the telecommunications industry. Each method has a different cost structure and hence will result in different equilibrium outcomes in general. For this reason, a welfare analysis of number portability needs to begin with a clear understanding of the costs associated with each technological choice. Thus, the first task of this paper is to describe the major technologies for providing telephone number portability, and to high-

[1] Perhaps predictably, this approach has not yet resulted in any firm commitment to introduce number portability. At the time of writing, the industry is preparing to commission a study into the net benefits of portability by different methods.

light the costs associated with each. This is addressed in Section 2 where we show that the per unit costs of completing calls to ported numbers is always higher than to non-ported numbers, and that some porting technologies raise the unit cost of all calls including those to non-ported numbers.

With this cost information as a background, in Section 3 we construct a model of competition between two suppliers of differentiated products in which the firms set two part tariffs. We model number portability through the use of a switching cost which is incurred by former customers of the incumbent subscribing to the entrant's service. In the absence of number portability, this switching cost is relatively high as consumers wanting to take the entrant's service must purchase new stocks of complementary goods (stationary, directory listings and, for business customers, advertising) that are dedicated to their old telephone number. The introduction of number portability has two impacts on the social costs of taking the entrant's service. First, it reduces the switching cost by substituting a lower cost network mediated technology which avoids the purchase of new complementary goods. Secondly, it increases the marginal cost of making calls because additional routing related tasks must be performed to establish connections.

It is useful to think of switching costs as having investment characteristics. In the absence of number portability, consumers considering taking the entrant's service are required to trade-off the investment in new complementary goods against the future stream of expected benefits from cheaper service. The incumbent has an incentive to exploit the hysteresis in this type of investment decision by setting high prices. If number portability increases marginal cost of production and thereby increases price of service, portability may not have quite the desired effect.

In our model, this trade off appear in the two components of the tariff.

Switching cost is reflected in the fixed fee of the tariff. Marginal cost of service production is reflected in the per unit price part of the tariff and thus the surplus which will be extracted by the fixed fee. The main part of the paper analyses a mature industry, where original cost of production is low enough that everyone buys from the incumbent prior to entry. Several cases are identified by their initial cost and demand characteristics. In the last section, we analyze an infant industry, where technology is too expensive for everyone to buy from incumbent prior to entry. In such a market, the analysis also depends on the fraction of the market covered prior to the entry.

An important conclusion of our work is that social gains are not automatically achieved, either for consumers or for society as a whole. Lower switching cost increases competition. However increase in marginal cost of production will make consumers worse. Depending on how the cost is allocated and exactly what proportion of consumers switch, the gain from competition may not be offset by the cost. Thus, number portability is not always and everywhere socially beneficial, a finding which contrasts markedly with the implicit regulatory assumption in a number of jurisdictions (Reinke, 1998).

Our work highlights the factors which are important determinants of policy outcomes. It therefore provides the basis for an empirical analysis of the likely overall impact of portability, and the incidence of the gains and losses. In this sense, our work can be seen as a component of optimal policy analysis in this area.

We adopt Laffont, Rey, and Tirole (1998) approach with differentiated products and heterogenous consumers to two-part tariff. Our analysis without switching cost is a special case of Calem and Spulber (1984) but we are able provide a more complete characterization of the prices and the welfare implications. It is not the goal of this paper to explain why competition does not change to linear pricing once entry occurs. We take as given the fact that competition in former monopoly network industries are in non-linear pricing. As we will see, a consumer can incorporate switching cost into purchasing decision naturally when prices are two-part tariffs.

Since Klemperer (1987) demonstrated effectiveness of switching cost as entry deterrent and their anti competitive nature (Klemperer (1987)) study of switching cost has focused on how firms use them strategically (Klemperer (1988,1995)). Farrell and Shapiro (1988) focused on switching costs generating network effects and Caminal and Matutes (1990) showed how switching cost can arise endogenously. There have also been studies on how firms counter switching cost of rivals (Fudenberg and Tirole (1997,1998), Chen (1997) through the price mechanism. Our focus is the welfare implications when switching cost is lowered technologically by a regulator. The only other paper that deals with regulatory reduction of switching cost (Gans and King (1999)) compare different methods of allocating the per unit switching cost to consumers with inelastic unitary demand. We consider different cost reduction technologies, with different implementations costs.

2 The Costs of Number Portability

In a standard set of interconnected telecommunications networks, the number dialled contains all information required for its successful routeing. This is no longer true once a number has been ported however, when two extra pieces of information are required. The first is the recognition that the number dialled has been ported to another network. The second is the most up to date routeing information. Porting technologies are distinguished by the way in which this extra information is supplied and used. It is useful to consider three alternatives.

2.1 Call Forwarding

The easiest way to provide number portability is to use the existing facilities for forwarding calls to a new number. Many carriers offer this technology as an add-on service to their existing customers, allowing them to temporarily route calls to a mobile phone, for example. Portability by means of call-forwarding can be implemented almost immediately and at minimal cost in most telephone networks. New investments in network infrastructure are avoided and the software required is already available. Basic call forwarding (CF) therefore provides an obvious starting point for our analysis.

Under a CF technology, calls to a ported number are initially directed to the exchange which hosted the B party prior to porting. This exchange, recognising that the B party has migrated to a different network, sets up a new circuit to the point of interconnection and then to the B party via the recipient carrier's network. In the process, two variable costs are incurred: one related to the call set-up which includes the provision of new routeing information; and the other being the cost of holding open more circuits than are minimally necessary to complete the call. The second of these costs occurs when the CF set-up path includes a "trombone" which is a circuit beginning at a trunking layer of the network, passing down to the original local exchange, and then back to the trunking layer.

In some networks it is possible to avoid tromboning a call by releasing the call path back to the trunk layer using a technology known variously as "drop-back", "crank-back" or "release-to-pivot". If this is possible, or can be made so with capital investment, then CF the only additional variable cost for calls to CF ported numbers is in the call set-up phase.

The additional fixed costs of CF porting are very low. These include entering new routing information into a database at the original local exchange, and the cost of any investment associated with achieving drop-back. It is likely, however, that the variable costs of CF will increase with the number of calls to ported numbers. This could occur as a result of congestion in the signalling network, and will be complicated by customers churning between several carriers. Consequently, CF is generally seen as being a useful initial technology, but one that may need to be replaced if and when porting rates increase.

2.2 Terminating IN

Intelligent network (IN) technologies involve the construction and maintenance of off-switch databases which map ported numbers to their new addresses. When a ported number is dialled, the database is consulted to obtain the new addressing data. The different IN technologies are distinguished through the location of the databases. Since the database must be located somewhere between the caller's local exchange and the B party's initial local exchange, we can usefully procede by taking each of these locations as defining one end of the spectrum of IN possibilities. We will refer to a system in which the database is located at (or near) the caller's local exchange an *originating* IN; and if the B party's initial local exchange (or a nearby trunk exchange) holds the database we call it a *terminating* IN.

If a terminating IN is installed for the purpose of achieving number portability, some fixed costs are incurred; these are associated with the hardware and software that comprises the IN. In addition, each time a customer ports their number, a once only cost of updating the IN database is incurred. Finally, calls to ported numbers impose variable costs associated with database look-ups and call re-routeing. The efficiency of the call routeing depends on exactly where the terminating IN database is located, and on whether drop-back is technically feasible. If the database is sited on the trunking layer, trombones to the

B party's initial local exchange are always avoided; this will also economise on the number of databases required, but each database will need more capacity and will be used more intensively. Given the specifications of the installed IN, however, some known mix of fixed and variable costs will arise. This mix differs in one important respect from those of an originating IN, to which we now turn.

2.3 Originating IN

The fixed costs associated with installing an originating IN are thought to be much greater than for a terminating IN, at least initially. This is because, with an originating IN that all exchanges need to have access to a database as soon as one number is ported. If the intelligence was located at the B party's end, as in the terminating IN, then databases are only needed in those geographical areas where the porting service is actually taken up.

A similar, but much more important, distinction occurs in respect of the variable costs of terminating and originating INs. Once a single number is ported, any given call *might* be directed at a ported number. Consequently, the IN database must be consulted for routeing information for *all* calls. This means that the cost of completing all calls increases, irrespective of whether the called number has been ported or not.

We have no information about the size of the costs attending each method of porting. It is, however, clear that each method imposes a different mix of fixed and variable costs, and that these may even generate cost externalities for calls to unported numbers.

Given cost information, and predictions about the proportions of calls that are directed to ported numbers, the cost minimising technology for implementing portability can be determined. These are the costs of achieving any welfare gains from introducing portability. As such, they feed directly into the welfare analysis conducted below. We now introduce a simple, but accurate model of the form of competition that number portability is thought to enhance.

3 Analysis

We analyse the implications of entry by a single entrant into a previously monopolised industry. We capture consumer switching costs, which are reduced by number portability, using a positive constant S. The cost of implementing a reduction in S, i.e., greater portability, will be reflected in the marginal costs of production for incumbent and entrant. We do not assume any relationship between the three parameters given various possibilities outlined in the previous section.

3.1 The Model

We consider a model of horizontal product differentiation. Consumers differ only by their preference over the products. We denote by T, the price of a product which consists of a fixed fee (F) and per unit price (p), i.e., $T = (p, F)$. Each consumer will buy one product only but may buy any number of units of the chosen product. The consumption choice is made in the following way: given the prices (T's) of products, a consumer calculates the optimal consumption choice for each product, compares the optimal consumption utility levels and buys the product that offers the highest utility level.

Consumers are distributed uniformly on the unit interval but are otherwise identical. Firm 0 is located at 0 and firm 1 at the other end, 1. Products are differentiated by the location of the seller.

Let $v(p) - F$ denote the indirect utility that a consumer achieves by consuming his most preferred product at price T. If the product is sold by firm 0, the indirect utility of consumer with preference (location x) will be $V_0(T, x) =$

$v(p) - F - tx$, where t is the "transportation cost" incurred by the consumer. The indirect utility of the same consumer of buying from firm 1 is $V_1(T,x) = v(p) - F - t(1-x)$. A consumer located at x will buy from firm 0 if and only if $V_0(T,x) \geq V_1(T,x)$ and $V_0(T,x) \geq 0$. A consumer may buy from neither firm, in which case they receive 0 indirect utility.

If firm 0 were a monopolist, consumers in the interval $[0, x(T)]$ will buy from firm 0 where $V_0(T, x(T)) = 0$ and $x(T) \in [0,1]$. Demand for firm 0's product will be $q(p)x(T)$. If $x(T) < 0$, the product is too expensive and no one buys. If $x(T) > 0$, everyone will buy and total demand will be $q(p)$. Given the marginal cost of production c, the monopolist sets T to maximize $\pi_m(T)$, where

$$\pi_m(T) = \begin{cases} 0 & \text{if } x(T) = \frac{v(p)-F}{t} < 0, \\ \{(p-c)q(p) + F\}x(T) & \text{if } 0 \geq x(T) \geq 1, \\ (p-c)q(p) + F & \text{if } x(T) < 1. \end{cases}$$

If we define $W = v(p) - F$, optimizing with respect to p and F is equivalent to optimizing with respect to p and W. The choice by the firm of how much surplus to extract (F) is equivalent to choosing how much to leave (W). Thus by "a price T" we mean both (p, F) and (p, W).

$$\pi_m(T) = \begin{cases} 0 & \text{if } x(T) = \frac{W}{t} < 0, \\ \{(p-c_0)q(p) + v(p) - W\}\, x(T) & \text{if } 0 \geq x(T) \geq 1, \\ (p-c_0)q(p) + v(p) - W & \text{if } x(T) > 1. \end{cases}$$

The surplus for consumers will be W less the transportation cost.

It is easy to show the following,

Lemma 1. *The solution $T^m(p^m, W^m)$ to the monopolist's problem (3.1) is*

(M) *If* $v(c_0) \geq 2t$*, then* $p^m = c_0$ *and* $W^m = t$.

(I) *Otherwise* $p^m = c_0$ *and* $W^m = v(c_0)/2$.

In case(M), marginal cost is low relative to the transportation cost so that all consumers will buy. We will call this a Mature Industry since everyone already buys from the incumbent monopolist prior to entry. This would be the case if the marginal cost of technology were very cheap relative to demand, and probably characterises the telephony industry in most developed economies. Another possibility is that the regulatory regime includes some type of subsidy (e.g. on output or input prices) to guarantee universal service via the price mechanism. In a mature industry, there is no benefit of having greater coverage as a result of another firm entering the market. Any benefits come from the competition that the entrant provides.

In case (I), the marginal consumer is $x(T^m) = \frac{v(c_0)}{2t} < 1$ and thus not everyone buys from the incumbent monopolist. This occurs because relative to the transportation cost t, the marginal cost of production c_0 is high, so that there are consumers who find it too costly to buy. We will refer to this case as Infant Industry.

In both cases, there is marginal cost pricing and the farthest buying consumer gets zero surplus.

3.2 Equilibrium of a Mature Industry after Entry

Firm 0 is the incumbent and firm 1 is the entrant. We assume that consumers who previously bought from the incumbent but now buy service from the entrant, incur a once-only switching cost S. Given the Mature Industry assumption, this switching cost applies to all of the entrant's customers. As noted above, S should strictly be viewed as an investment from which customers expect to receive a flow of benefits in the future. Without loss of generality, we abstract from these dynamic considerations and regard S as the share of the actual switching cost which would be apportioned to the current period if this total cost were amortised over the expected natural life of the service contract [2].

Let the price $T_i = (p_i, F_i)$ denote the price offered by firm i. We consider a game with the two firms as the players in which the prices T_i are chosen simultaneously, and the profits are the payoffs. We now characterize the Nash equilibrium of this game.

The utility from buying from firm 1 (the entrant) will be $V_1(T, x) = v(p) - F - t(1-x) - S$. Hence a consumer located at x will buy from firm 0 iff $V_0(T_0, x) \geq V_1(T_1, x)$ and $V_0(T_0, x) \geq 0$ and similarly for buying from firm 1. We can therefore define the bench marks, $\hat{x}_0(T_0)$, $\hat{x}_1(T_1)$, and $\hat{x}(T_0, T_1)$, by,

$$V_0(T_0, \hat{x}_0(T_0)) = 0, \quad V_1(T_1, \hat{x}_1(T_1)) = 0, \tag{1}$$

$$V_0(T_0, \hat{x}(T_0, T_1)) = V_1(T_1, \hat{x}(T_0, T_1)). \tag{2}$$

All consumers to left (right) of $\hat{x}_0(T_0)$ ($\hat{x}_1(T_1)$) have positive utility buying from firm 0 (firm 1). All consumers to left (right) of $\hat{x}(T_0, T_1)$ have greater utility from buying from firm 0 (firm 1). By definition, it must be that either (i) $\hat{x}_0(T_0) < \hat{x}(T_0, T_1) < \hat{x}_1(T_1)$, or (ii) $\hat{x}_0(T_0) \geq \hat{x}(T_0, T_1) \geq \hat{x}_1(T_1)$. In case (i), there is an interval of consumers in the middle that do not buy at all. In case (ii), all consumers will buy with three possibilities: all buy from firm 0 if $\hat{x}(T_0, T_1) \leq 0$, all buy from firm 1 if $\hat{x}(T_0, T_1) \geq 1$, and otherwise there is positive sales by both firms.

Again, we use the surplus W_i as a choice variable instead of F_i. Specifically,

[2] In the absence of geographic portability, physical relocations will generally require new numbers to be issued. Hence we can usefully think of the cost incurred by a consumer in changing numbers being spread across the number of billing periods expected in the current location.

$W_0 = v(p) - F_0$ and $W_1 = v(p) - F_1 - S$. Since W_1 already takes into account the switching cost when buying from firm 1, W_1 only needs to cover the transportation cost. From this substitution, we have $\hat{x}_0(T_0) = \frac{W_0}{t}$, $1 - \hat{x}_1(T_1) = \frac{W_1}{t}$, and $\hat{x}(T_0, T_1) = \frac{W_0 - W_1 + t}{2t}$.

Firm 0's profit is a function of (p_0, W_0) and (p_1, W_1):

$$\pi_0 = \begin{cases} \pi_0^A = \{(p_0 - c_o)q(p_0) + v(p_0) - W_0\} \frac{W_0}{t} & \text{for } W_0 \leq t - W_1, \\ \pi_0^B = \{(p_0 - c_0)q(p_0) + v(p_0) - W_0\} \frac{W_0 - W_1 + t}{2t} & \text{for } t - W_1 < W_0 \leq t + W_1, \\ \pi_0^C = (p_0 - c_0)q(p_0) + v(p_0) - W_0 & \text{for } W_1 + t < W_0. \end{cases}$$

It is easy to show using Roy's Identity that for any W_0 and (p_1, W_1), all segments of the function are maximized with respect to p_0 at $p_0 = c_0$, i.e., marginal cost pricing. This maximises the indirect utility of every consumer. Now the problem is to find the W_0 to maximize,

$$\pi_0 = \begin{cases} \pi_0^A = \{v(c_0) - W_0\} \frac{W_0}{t} & \text{for } W_0 \leq t - W_1, \\ \pi_0^B = \{v(c_0) - W_0\} \frac{W_0 - W_1 + t}{2t} & \text{for } t - W_1 < W_0 \leq t + W_1, \\ \pi_0^C = v(c_0) - W_0 & \text{for } W_1 + t < W_0. \end{cases}$$

Note that the problem is independent of p_1. Straightforward but tedious calculation yields the following.

Lemma 2. *Firm 0's best response correspondence $W_0 = R_0(W_1)$ is,*

(1) If $t < \frac{v(c_0)}{3}$, then

$$R_0(W_1) = \begin{cases} t + W_1 & \text{for } W_1 \leq v(c_0) - 3t, \\ \frac{v(c_0) + W_1 - t}{2} & \text{for } W_1 \geq v(c_0) - 3t. \end{cases}.$$

(2) If $t > \frac{v(c_0)}{3}$*, then*

$$R_0(W_1) = \begin{cases} t - W_1 & \text{for } W_1 \leq t - \frac{v(c_0)}{3}, \\ \frac{v(c_0)+W_1-t}{2} & \text{for } t - \frac{v(c_0)}{3} \leq W_1. \end{cases}.$$

(3) If $t = \frac{v(c_0)}{3}$*, then*

$$R_0(W_1) = \frac{v(c_0) + W_1 - t}{2} \textit{for all } W_1 \geq 0.$$

Firm 1's best response correspondence is obtained similarly, and differs only by the fact that the switching cost must be taken into account in the profit function. Using the same argument as with firm 0, firm 1 chooses W_1 to maximize,

$$\pi_1 = \begin{cases} \pi_1^A = (v(c_1) - W_1 - S)\frac{W_1}{t} & \text{for } W_1 \leq t - W_0, \\ \pi_1^B = (v(c_1) - W_1 - S)\frac{t-W_0+W_1}{2t} & \text{for } t - W_0 < W_1 \leq t + W_0, \\ \pi_1^C = v(c_1) - W_1 - S & \text{for } t + W_0 < W_1. \end{cases}$$

Lemma 3. *Firm 1's best response correspondence* $W_1 = R_1(W_0)$ *is,*

(1) If $t < \frac{v(c_1)-S}{3}$*, then*

$$R_1(W_0) = \begin{cases} \frac{v(c_1)-S}{2} \text{ or } t + W_0 & \text{for } W_0 \leq t - \frac{v(c_1)-S}{2}, \\ t + W_0 & \text{for } t - \frac{v(c_1)-S}{2} < W_0 \leq v(c_1) - 3t, \\ \frac{v(c_1)+W_0-t}{2} & \text{for } v(c_1) - S - 3t < W_0. \end{cases} \cdot$$

(2) If $t > \frac{v(c_1)-S}{3}$*, then*

$$R_1(W_0) = \begin{cases} \frac{v(c_1)-S}{2} & \text{for } W_0 \leq t - \frac{v(c_1)-S}{2} \\ t - W_0 & \text{for } t - \frac{v(c_1)-S}{2} < W_0 \leq t - \frac{v(c_0)}{3}, \\ \frac{v(c_1)+W_0-t}{2} & \text{for } t - \frac{v(c_1)-S}{3} \leq W_0. \end{cases} .$$

(3) If $t = \frac{v(c_1)-S}{3}$*, then*

$$R_1(W_0) = \frac{v(c_1) - S + W_0 - t}{2} \textit{for all } W_0 \geq 0.$$

In case (1), the value of $R_1(W_0)$ for $W_0 \leq t - \frac{v(c_1)-S}{2}$ is $\frac{v(c_1)-S}{2}$ if $\pi_1^A(\frac{v(c_1)-S}{2}) \geq \pi_1^B(t + W_0)$ and the value is $t + W_0$ otherwise. It will always be the case that $R_1(W_0) > W_0$ which guarantees that this segment of the best response function never contains the Nash equilibrium (in pure strategies). Because of the switching cost, firm 1 may not always want to sell to all buyers not buying from firm 0. However, because of the Mature Industry assumption, firm 0 will never want to miss making a sale to a buyer who does not buy from firm 1. Using the best-response correspondences, we can characterize the Nash equilibrium prices and allocations.

For both firms, there is a case (case (2) for both) for which strategies can be strategic complements. Competition in fixed fees is effectively competition in prices which are strategic substitutes: when the rival firm lowers its fee, a firm's optimal response is to also lower its fee. That is, when rival increases demand, each firm finds it profitable to reduce its fee and increase demand (to take back some of the loss in demand due to the rival's fee decrease). In doing so, each firm must forego some surplus it previously collected from its captive consumers. In case (2) however if $W_1 \leq t - \frac{v(c_0)}{3}$, then in response to rival fee decrease, firm 0 finds it optimal to increase its own fee (and further give up demand) to extract more surplus from its captive consumers. For this to be optimal, the reduction in demand due to fee increase must be small relative to surplus, i.e., transportation cost (t) must be sufficiently large, which is the condition for case (2). In addition, the marginal consumer's surplus must be small so that it is not worth retaining ($W_1 \leq t - \frac{v(c_0)}{3}$). A similar argument holds for firm 1's strategic complementarity.

Proposition 1. *Post entry equilibrium with switching costs ($S > 0$) for a Mature Industry ($v(c_0) \geq 2t$) is characterised by marginal cost pricing, $p_i^* = c_i$ in all regimes identified below. The equilibrium fixed fees and outcomes for each regime are detailed below.*

(I) If $v(c_1) - S \geq v(c_0) + 3t$, the equilibrium prices are

$$W_0^* = v(c_1) - S + t, \quad W_1^* = v(c_1) - S.$$

In this case, all consumers buy from the incumbent. The equilibrium profits are,

$$\pi_0^* = v(c_0) - (v(c_1) - S) - t, \quad \pi_1^* = 0.$$

(II) If $v(c_1) - S \leq v(c_0) - 3t$, then equilibrium prices are

$$W_0^* = v(c_0), \quad W_1^* = v(c_0) + t.$$

Now, all consumers buy from the entrant. The equilibrium profits are,

$$\pi_0^* = 0, \quad \pi_1^* = v(c_1) - S - v(c_0) - t.$$

(III) If $v(c_0) + v(c_1) - S \geq 3t$ and $v(c_0) - 3t < v(c_1) - S < v(c_0) + 3t$, then

equilibrium prices are

$$W_0^* = \frac{v(c_1) - S + 2v(c_0) - 3t}{3}, \quad W_1^* = \frac{2(v(c_1) - S) + v(c_0) - 3t}{3}.$$

Both firms make positive sales. The marginal consumer is at $\hat{x}(T_0^*, T_1^*) = \frac{1}{2} + \frac{v(c_0)-v(c_1)+S}{6t}$ *and has positive surplus* $\frac{v(c_0)+v(c_1)-S-3t}{2}$*. The equilibrium profits are,*

$$\pi_0^* = \frac{1}{2t}\left\{\frac{v(c_0) - (v(c_1) - S)}{3} + t\right\}^2, \quad \pi_1^* = \frac{1}{2t}\left\{\frac{v(c_1) - S - v(c_0)}{3} + t\right\}^2.$$

(IV) If $v(c_0) + v(c_1) - S < 3t$*, then there are continuum of equilibria. The equilibrium prices indexed by* $\alpha \in [0, 1]$ *are,*

$$W_0^* = \alpha\frac{v(c_0)}{3} + (1-\alpha)(t - \frac{v(c_1) - S}{3}), \quad W_1^* = \alpha(t - \frac{v(c_0)}{3}) + (1-\alpha)\frac{v(c_1) - S}{3}.$$

The marginal consumer is at $\hat{x}(T_0^*, T_1^*) = \hat{x}_0(T_0^*) = \hat{x}_1(T_1^*) = \alpha\frac{v(c_0)}{3t} + (1-\alpha)(1 - \frac{v(c_1)-S}{3t})$ *and has zero surplus. The equilibrium profits are,*

$$\pi_0^* = \frac{W_0^*}{t}(v(c_0) - W_0^*), \quad \pi_1^* = \frac{W_1^*}{t}(v(c_1) - S - W_1^*).$$

The best-response correspondences and equilibria for regimes (III) and (IV) are demonstrated in Figure 1. Regime (I) is when $\frac{v(c_0)-t}{2}$ is large relative to $\frac{v(c_1)-S-t}{2}$ so that equilibrium occurs on the $W_1 = W_0 - t$ segment of R_0. This occurs either when entrant is significantly less efficient, or when the switching cost is very large, or both. Entry will not result in any consumers actually switching to the new supplier. However, because of the entrant, the consumers have higher surplus. In particular, the surplus of the consumer at $x = 1$ increases from 0 when the incumbent was a monopolist to W_1^* after entry. $x = 1$ is the marginal consumer and so is exactly indifferent between switching and not switching.

Regime (II) is when $\frac{v(c_0)-t}{2}$ is small relative to $\frac{v(c_1)-S-t}{2}$. In this case, the equilibrium occurs on the $W_1 = W_0 + t$ segment of R_1. This will occur when the entrant is very efficient and the switching cost is low enough so that all consumers switch. Again, the entrant's fixed fee is constrained by the option consumers have of not switching. The consumer at $x = 0$ has positive surplus of W_0^*.

Both firms have positive sales in regimes (III) and (IV). Firms equally split the market when $v(c_0) = v(c_1) - S$, which is a subregime of regime (III). Because of the switching cost, the entrant must be more efficient in order to have the same market share. The entrant will not reduce the final surplus by the whole amount of the switching cost because it takes into account the fact that the incumbent will also reduce its surplus in response. This is direct result of strategic complementarity. For both groups of consumers, equilibrium surplus decreases with the switching cost. But the equilibrium fee only increases for the incumbent. The entrant's fee decreases because of the switching cost (switching cost is the "wedge"),

$$F_0^* = \frac{v(c_0) - (v(c_1) - S) + 3t}{3} \quad F_1^* = \frac{v(c_1) - S - v(c_0)) + 3t}{3}.$$

The incumbent charges a higher fee and increases its market share with higher switching cost so its profit is increasing in switching cost. The entrant has a lower market share and a lower fee so its profit decreases with the switching cost.

In regime (IV), the intersection of the best-response correspondences is the closed line segment between points $(\frac{v(c_0)}{3}, t - \frac{v(c_0)}{3})$ and $(t - \frac{v(c_1)-S}{3}, \frac{v(c_1)-S}{3})$. Among these equilibria, the one with largest share for the incumbent, $W_0^* = t - \frac{v(c_1)-S}{3}$ is the most profitable for the incumbent. This corresponds to $\alpha = 0$

in the proposition.

It is interesting to note that this equilibrium coincides with the subgame perfect equilibrium outcome if prices were determined sequentially and the incumbent chooses first. This is because the best-response correspondence of the entrant (the second mover) is kinked at this point. Prices change from strategic substitutes to strategic complements at this point. The equilibrium reflects the strategic substitute nature of the strategies. When the switching cost increases, the surplus of entrant buyers increases while that of entrant decreases. The equilibrium fees for both decrease with higher switching cost.

$$F_0^* = \frac{3-\alpha}{3}v(c_0) - (1-\alpha)(t - \frac{v(c_1)-S}{3}), \quad F_1^* = (\frac{2+\alpha}{3}(v(c_1)-S) - \alpha(t - \frac{v(c_0)}{3}).$$

When switching costs increases, the entrant's equilibrium share decreases and the fee is lower. So the entrant's profit decreases with switching cost. Higher switching costs result in lower fees but greater market share for the incumbent. Thus if the fee is relatively large profits are increasing in switching cost.

If, in addition to the Mature Industry assumption, we also assume that the entrant is sufficiently efficient, i.e., $v(c_1) - S \geq 2t$, then regime (IV) will never occur and equilibrium will always be unique.

Corollary 1. *The corresponding consumer surplus for the four regimes are,*

(I) $CS_I = \frac{1}{2}(2W_0^* - t)$,

(II) $CS_{II} = \frac{1}{2}(2W_1^* - t)$,

(III) $CS_{III} = (W_0^* + W_1^* - t(1-\hat{x}(T_0^*, T_1^*)))\frac{\hat{x}(T_0^*, T_1^*)}{2} + (W_1^* + W_0^* - t\hat{x}(T_0^*, T_1^*))\frac{1-\hat{x}(T_0^*, T_1^*)}{2}$
$= \frac{1}{2}(W_0^* + W_1^*) - t\hat{x}(T_0^*, T_1^*)(1-\hat{x}(T_0^*, T_1^*)$,

(IV) $CS_{IV} = \frac{1}{2}\left\{W_0^*\frac{W_0^*}{t} + W_1^*\left(1 - \frac{W_0^*}{t}\right)\right\} = \frac{1}{2t}\{(W_0^*)^2 + (W_1^*)^2\}$.

When the switching cost is reduced, part of R_1 moves upward. Increases in marginal cost of production c_i move part of R_i downward. Within regime (III), a reduction in S will unambiguously increase consumer surplus. However if this results in S being higher than c_i, the total effect might be to reduce consumer surplus. This is because the equilibrium might then change from one regime to another by parameter changes. As a result, it is more useful to analyze welfare in the space of $v(c_0)$ and $v(c_1) - S$.

3.2.1 Welfare Analysis of Number Portability in a Mature Industry

Using the proposition and the corollary, we can find the equilibrium consumer surplus and producer surplus as functions of the parameters of the model. In regime (IV) where there are multiple equilibria, we choose the one that yields the highest payoff for the incumbent ($\alpha = 0$).

The equilibrium consumer surplus and producer surplus for each of the four regimes (defined in Proposition 1) are summarized below. The iso-consumer surplus lines are shown in Figure 2.

	Consumer Surplus	Producer Surplus
I	$v(c_1) - S + \frac{t}{2}$	$v(c_0) - (v(c_1) - S) - t$
II	$v(c_0) + \frac{t}{2}$	$v(c_1) - S - v(c_0) - t$
III	$\frac{((v(c_0)-v(c_1)+S))^2}{36t} + \frac{v(c_1)-S+v(c_0)}{2} - \frac{5}{4}t$	$\frac{1}{9t}\left(v(c_0) - v(c_1) + S)\right)^2 + t$
IV	$\frac{1}{2t}\left\{(t - \frac{v(c_1)-S}{3})^2 + (\frac{v(c_1)-S}{3})^2\right\}$	$v(c_0) - t - \frac{(v(c_0)-2t)(v(c_1)-S)}{3t} + \frac{(v(c_1)-S)^2}{9t}$

Table 1: Consumer and Producer Surplus by Regime

In both regimes (I) and (II), the consumers are served by only one of the firms. In regime (I), the incumbent is the sole supplier and thus consumers never actually switch. However the fee they pay reflects the cost of switching: the higher the cost, the greater the fee that the incumbent can charge. Thus consumer surplus is decreasing in the switching cost. The switching cost is actually "collected" by the incumbent and thus its profit is increasing in switching cost. The sum of the two surpluses, however, does not depend on switching cost, so the switching cost determines the split of the surplus between consumers and the incumbent. Recall that entry (although entrant does not actually sell anything) increases consumer surplus. So switching cost will syphon some of the benefit to the incumbent.

Since everyone switches in equilibrium in regime (II), switching costs are actually incurred. Thus consumer surplus is increasing in the switching cost but producer surplus is independent of the switching cost because the fee, and thus size of demand in equilibrium, is independent of switching cost.

In regime (III), the switching cost is anti-competitive in the standard sense: consumer surplus decreases and producer surplus increases in the switching cost. This is because the switching cost reduces temptation to cut prices and therefore decreases surplus of consumers. In addition to reducing competition, the switching cost is a cost paid but never collected by anyone within this model.[3] This also contributes to social surplus reduction with higher switching cost.

In regime (IV), the switching cost increases consumer surplus but it is questionable if this is actually pro-competitive. In this regime, higher switching costs increase the surplus for each of the consumers buying from the incumbent and reduce the surplus for those buying from entrant. In addition, the proportion of those buying from the incumbent increases. This increases total consumer surplus. Higher switching costs increase consumer welfare by skewing the surplus distribution so that there are more people in the higher surplus consumer group (which is advantaged) and fewer in the lower surplus consumer group (which is

[3] More generally, the switching cost without number portability is higher than it need be.

disadvantaged). Producer surplus decreases for a similar reason.

The iso-social surplus curves are presented in Figure 3. Given same level of switching cost reduction, technology that changes marginal cost may be better. By allocating the cost accordingly, it may be possible to realize distributional gain at the same time. For instance, unless S is reduced to zero, technology that increases c_0 more than c_1 may achieve $v(c_1) - S = v(c_0)$.

Recall that in regime (III), consumer surplus was always decreasing in S. Social surplus may increase with S in some regions of (III) if producer surplus gain is large enough. This occurs where $v(c_1) - S \geq \frac{9}{5}t + v(c_0)$ or $v(c_1) - S \leq -\frac{9}{5}t + v(c_0)$. In these regions one firm is significantly more efficient implying significant market power of one of the firms. In this case increasing switching cost doest not hurt consumers that much at the margin while producers gain significantly. Given our assumption $v(c_1) \leq v(c_0)$, S reduction should always be done with a technology that increases c_0, put all the burden on the incumbent if S is very large (and is in this region).

3.3 Analysis of an Infant Industry

In an infant industry, the incumbent will only sell to $x(T^m) = \frac{v(c_0)}{2t} < 1$ of the market as a monopolist. Not all potential buyers of service buy, which might be the case for the mobile phone market. Thus if the incumbent behaved myopically as a static monopolist prior to entry, there will be $1 - x(T^m)$ of consumers who would not incur the switching cost in order to buy from the entrant once entry occurs. It is possible for the incumbent to impose switching costs on some additional consumers by selling beyond $x(T^m)$ prior to entry. The monopolist could optimize dynamically by trading off static losses prior to entry for a gain after entry (Klemperer (1995)).

To answer this question fully, we extend our model by adding another period, prior to entry, where incumbent is a monopolist. However we restrict the monopolist's prices to be independent over periods. That is, there are no two periods contracts available in the first period. In the second period, the incumbent and the entrant each choose a price simultaneously. The entrant is unable to discriminate between consumers that bought from incumbent and hence incur switching cost, and those that did not and therefore have no switching cost.[4]

We characterize the subgame perfect Nash equilibrium of this game. There will be one subgame for each of the prices that incumbent chooses in the first period. Since the prices are independent over periods, the only relevant fact for each subset is the proportion of the market the incumbent sold to in the first period, x^m. To characterize the subgame perfect equilibrium, we only need to characterize second period prices as function of x^m.[5]

3.3.1 Post Entry Equilibrium of an Infant Industry

In this section we analyze the (second period) post entry behaviour of the firms, given x^m. Again, firm 0 is the incumbent and firm 1 is the entrant. We now make the following transformation of the prices, (p_i, F_i),

$$W_0 = V_0(T_0, 0) = v(p_0) - F_0, \quad W_1 = V_1(T_1, 1) + S = v(p_1) - F_1.$$

The surplus from buying from the incumbent, W_0, is the same as before. The surplus W_1 is not always the final surplus (gross of transportation cost) because S has not been subtracted. For those consumers that bought from the incumbent previously ($x \leq x^m$), final surplus from buying from firm 1 will be $W_1 - S - t(1-x)$. For those that did not buy previously from the incumbent ($x > x^m$),

[4]This assumption is probably unrealistic but is unlikely to affect the aggregate analysis of consumer or social surplus. Depending on how the cost of porting services are allocated, however, price discrimination could result in large transfers from low value consumers to high value consumers who churn between carriers.

[5]Given independence of prices over the two periods, it is also easy to see that there will be marginal cost pricing in the first period. Thus x^m has one to one relationship with fixed fee.

final surplus will be $W_1 - t(1-x)$. Again, we refer to (p_i, W_i) as firm i's price.

We now define the marginal consumers, $\hat{x}_0(T_0)$, $\hat{x}_1(T_1)$, and $\hat{x}(T_0, T_1)$ by

$$W_0 - t\hat{x}_0(T_0) = 0, \quad W_1 - t(1 - \hat{x}_1(T_1)) = 0,$$
$$W_0 - t\hat{x}(T_0, T_1) = W_1 - t(1 - \hat{x}(T_0, T_1))$$

$\hat{x}(T_0)$ is the same as that defined by (1) but the other two differ since there is no switching cost.

We define $\hat{x}_1^S(T_1)$ and $\hat{x}^S(T_0, T_1)$, taking into account the switching cost:

$$W_1 - S - t(1 - \hat{x}_1^S(T_1)) = 0, \quad W_0 - t\hat{x}^S(T_0, T_1) = W_1 - S - t(1 - \hat{x}^S(T_0, T_1))$$

Depending on if x had previously bought from incumbent or not, the relevant surplus is $W_1 - S - t(1-x)$ or $W_1 - t(1-x)$.

Let the incumbent's pre-entry share be denoted x^m. Consider the case when incumbent's fee is very large post entry so that $W_0 < x^m t$. The entrant has a local monopoly over those that did not buy from incumbent previously. Let $W(x^m)$ be the surplus that makes the consumer at x^m exactly indifferent between buying from entrant and not buying. That is,

$$W(x^m) - t(1 - x^m) = 0.$$

For $W_1 < W(x^m)$, demand for the entrant's service (y_1) is $y_1 = 1 - \frac{W_1}{t}$. Increasing surplus beyond $W(x^m)$ by a very small amount will not increase demand for the entrant's service because now the entrant must induce consumers to switch. In fact demand for entrant's service will not increase from $y_1 = 1 - x^m$ unless W_1 is greater than $W(S)$ defined by

$$W(S) - S - t(1 - x^m) = 0.$$

For surplus greater than $W(S)$, demand for the entrant's service is $y_1 = 1 - \frac{W_1 - S}{t}$ until $W_1 - S - t(1-x) = W_0 - tx$. Then the consumer chooses between buying from the incumbent again or switching. Then for such W_1, demand is $y_1 = 1 - \hat{x}^S$.[6] See Figure 4.

When $W_0 > x^m t$, then entrant has local monopoly until $x_1 = x_0$. For larger W_1, demand is $y_1 = 1 - \hat{x}$. This is true until $W_1 - t(1-x^m) = W_0 - tx^m$. Then demand will be $y_1 = 1 - x^m$ and will remain at that level until $W_1 - S - t(1 - x^m) = W_0 - tx^m$. Beyond that, demand will be $y_1 = 1 - \hat{x}^S$. This is summarized in Figure 5 along with demand for incumbent's service, y_0.

Given these demand functions, we can derive the profit functions which depend on the prices and x^m. From the profit functions, we can again characterise the best-response correspondences. It is easy to show that, as before, there will be marginal cost pricing, $p_i^* = c_i$. Thus we only need to characterize the fixed fees, or W_i's.

Proposition 2. *If the first period share is sufficiently large and if the transportation cost and the switching cost is both large relative to marginal costs, then the incumbent's post entry share will be the same as first period share, and the entrant will sell to the rest of the market. Specifically, if $x^m \geq \frac{v(c_0)}{2t}$ and*

$$\frac{3t - v(c_1)}{4t} < x^m < \frac{3t - (v(c_1) - S)}{4t},$$

then the post entry equilibrium will be,

$$W_0^* = x^m t, \quad W_1^* = 1 - x^m t.$$

Because of the switching cost, the entrant does not find it profitable to reduce the fixed fee to get any of the consumers to switch. It charges enough to get the

[6] We omit the augments of functions $\hat{x}^S(T_0, T_1)$, $x_1(T_1)$, $x_0(T_0)$ since there is no confusion.

marginal costumer at x^m. Since x^m is greater than what the incumbent would sell to if it were a monopolist, there is no incentive for it to sell more. If we let $x^m = \frac{v(c_o)}{2t}$, we have the following.

Corollary 2. *If*

$$\frac{3t - v(c_1)}{2} < v(c_0) < \frac{3t - (v(c_1) - S)}{2},$$

then the incumbent will get the static monopoly profit in both periods in equilibrium. The entrant will sell to the rest of the market after entry.

Not only will the incumbent sell to the monopoly share of the market, it also gets the monopoly profit.

There will be gains in consumer surplus of course. These will come from the surplus of those consumers who did not buy previously. Notice, however, there will be no gain to consumer surplus as a direct result of newly introduced competition between the firms.

4 Conclusion

This analysis has demonstrated that the widespread presumption in favour of number portability is not necessarily in the interests of society in general, or even of consumers. In well developed telephony markets with high penetration rates, it is possible for consumers as a group to receive less surplus following a reduction in the cost of switching between carriers as a result of the introduction of number portability. We have examined four possible mature industry regimes, two of which involve some sharing of the market between the incumbent and the the entrant. Switching costs affect consumers and firms differently in these two shared market cases. In one case (regime (III)) consumers gain and firms lose when the switching cost falls, while in the other case (regime (IV)) these effects are reversed. In both of these cases, the net welfare effect of a policy that reduces the switching cost is theoretically ambiguous.

The situation is more clear cut for developing telephony markets with low initial penetration rates. In this case, under plausible assumptions, reductions in the cost of switching benefit consumers and the entrant, and have no effect on the incumbent. The consumer gains come entirely from expansion of the market.

It is important to realise that our model only analyses gross surplus gains. Consequently, even in cases where total surplus or consumer surplus is enhanced by lower switching costs, the net effect of portability could be negative if the fixed and variable costs of providing a porting service are sufficiently large.

Our analysis is therefore only the first step in a full welfare calculus, with more work being required to complete the task. Two empirical studies are required: a demand side analysis of consumer valuations; and information on both the costs of implementing portability, and the difference that this would make to current levels of switching costs.

One further implication of our work is that number portability is likely to result in transfers between individual consumers. An example of this is apparent in regime (IV) of the mature industry, where switching costs (and hence changes in this cost) served to skew the distribution of returns across consumers, advantaging some at the expense of others. In the event that porting customers do not bear the full cost of their actions, additional transfers are inevitable.

References

Calem, Paul S. and Spulber, Daniel F., 1984. "Multiproduct Two Part Tariffs." *International Journal of Industrial Organization* 2:105-115.

Caminal, Ramaon and Matutes, Carmen, 1990. "Endogenous swithcing costs in a duopoly model," *International Journal of Industrial Organizaiton"* 8:353-373.

Chen, Yongkin, 1997. "Paying customers to switch." *Journal of Economics & Management Strategy* 4:877-897.

Farrell, Joseph and Shapiro, Carl, 1988. "Dynamic competition with switching costs." *RAND Journal of Economics* 19(1):123-137.

Fudenberg, Drew and Tirole, Jean, 1997. "Customer Poaching and Brand Switching." Mimeo.

Gans, Joshua and King, Stephen, 1999. "Regulating Endogenous Customer Switching Costs." Mimeo, University of Melbourne.

Klemperer, Paul, 1987a. "The competitiveness of markets with switching costs." *Rand Journal of Economics* 18(1):138-150.

Klemperer, Paul, 1987b. "Entry deterrence in markets with consumer switching costs." *The Economic Journal* 97:99-117.

Klemperer, Paul, 1988. "Welfare effects of entry into markets with switching costs." *The Journal of Industrial Economics* 37(2):159-165.

Klemperer, Paul, 1995. "Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade." *Review of Economics Studies* 62:515-539.

Laffont, Jean-Jacques, Patrick Rey, and Jean Tirole, 1998. "Network competition: I Overview and nondiscriminatory Pricing." *RAND Journal of Economics* 29(1):1-37.

Oren, Shmuel S., Stephen A. Smith, and Robert B. Wilson, 1983. "Competitive

Nonlinear Tariffs." *Journal of Economic Theory* 29:49-71.

Reinke, Thomas H., 1998. "Local number portability and local loop competition." *Telecommunications Policy* 22(1):73-87.

Spulber, Daniel F. 1981. "Spatial nonlinear pricing." *American Economic Review* 71(5):923-33.

Appendix

Iso-social surplus curves are obtained from,

$$
\begin{aligned}
&(I)\ SS = CS + PS = -\frac{1}{2}t + v(c_0),\\
&(II)\ SS = -\frac{1}{2}t + v(c_1) - S.
\end{aligned}
$$

For the remaining regimes, using the following partial derivatives,

$$
\begin{aligned}
(III)\ \frac{\partial SS}{\partial v(c_0)} &= \frac{1}{2}t + \frac{5(v(c_0) - v(c_1) + S}{18t},\\
\frac{\partial SS}{\partial (v(c_1) - S)} &= \frac{1}{2}t - \frac{5(v(c_0) - v(c_1) + S}{18t},\\
(IV)\frac{\partial SS}{\partial v(c_0)} &= \frac{3t - v(c_1) + S}{3t},\\
\frac{\partial SS}{\partial (v(c_1) - S)} &= \frac{4(v(c_1) - S) - 3v(c_0) = 3t}{9t}.
\end{aligned}
$$

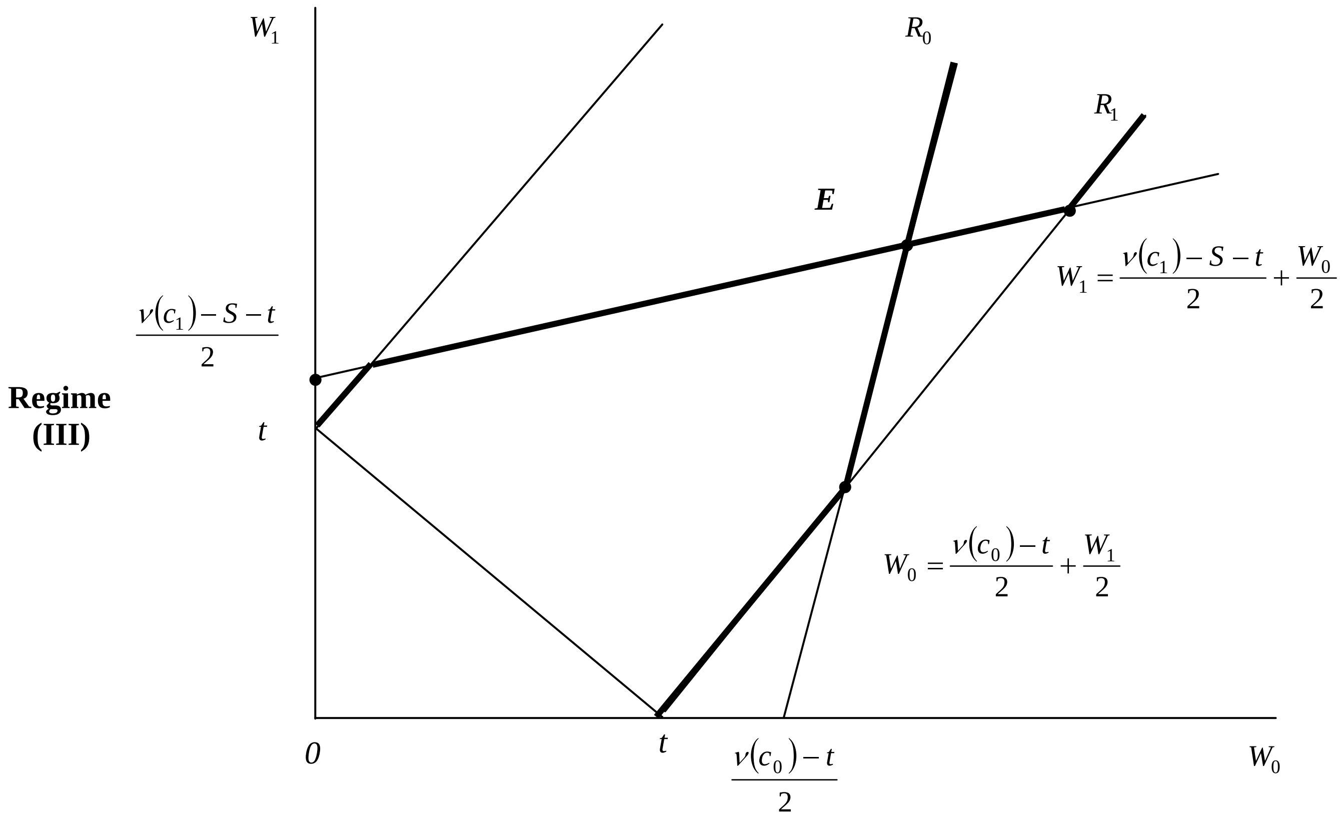


Figure 1a: Best Response Correspondences (Regime III)

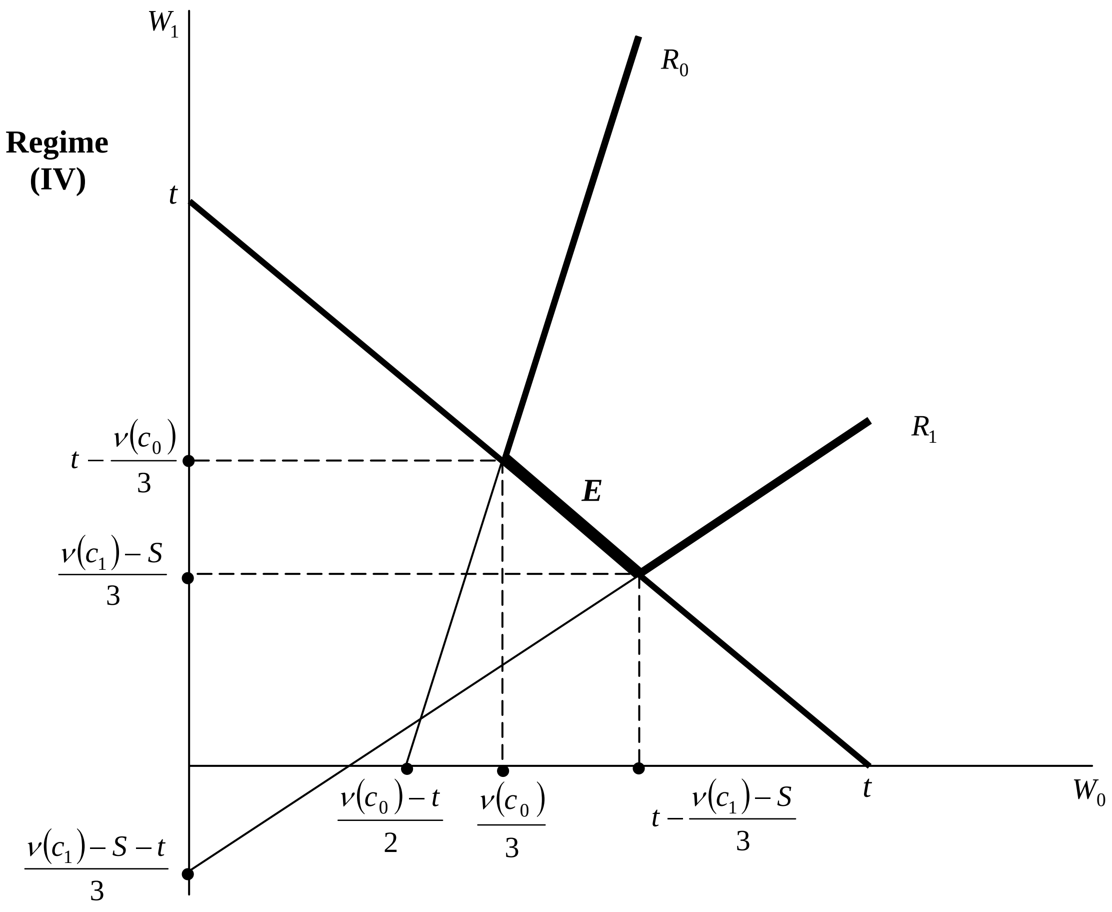


Figure 1b: Best Response Correspondences (Regime IV)

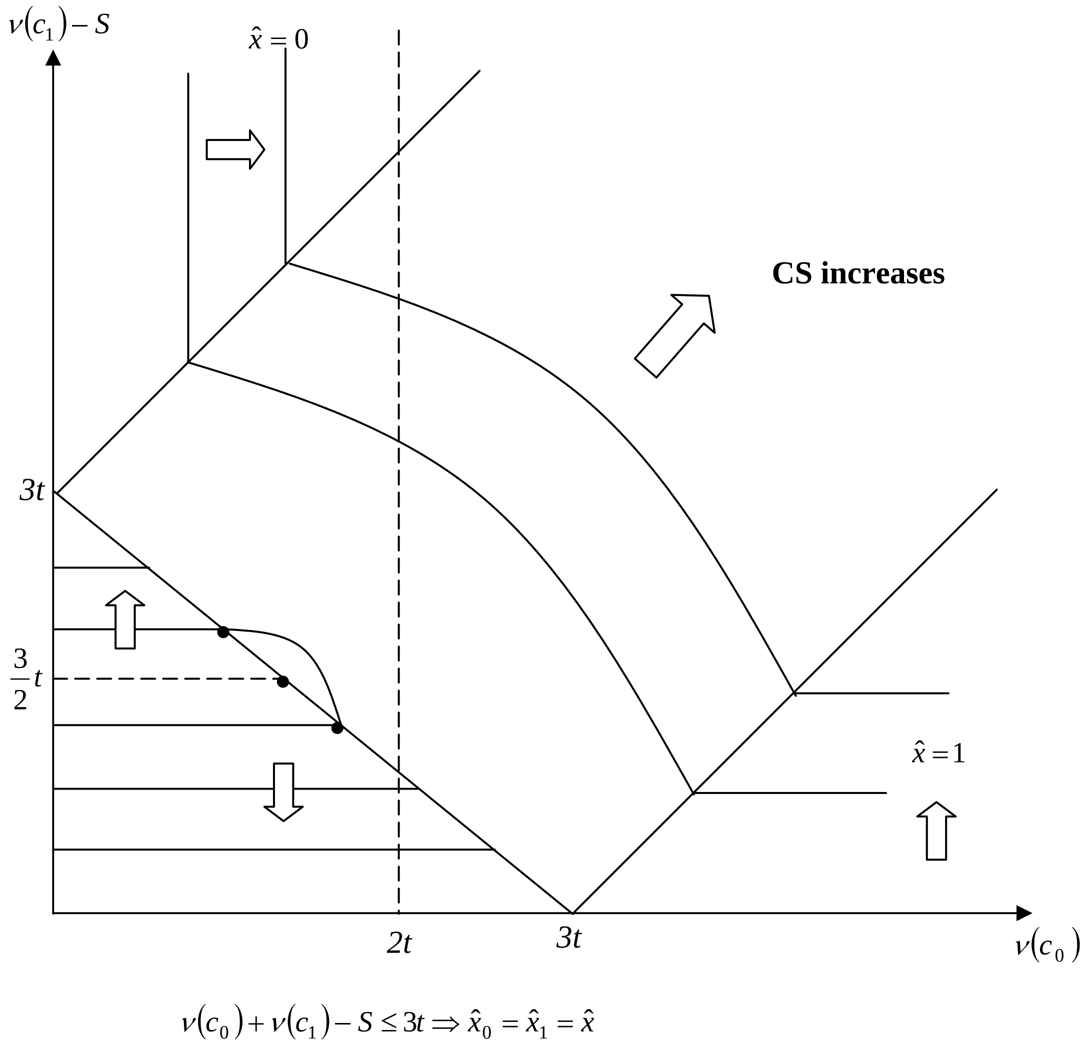


Figure 2: Iso Consumer Surplus

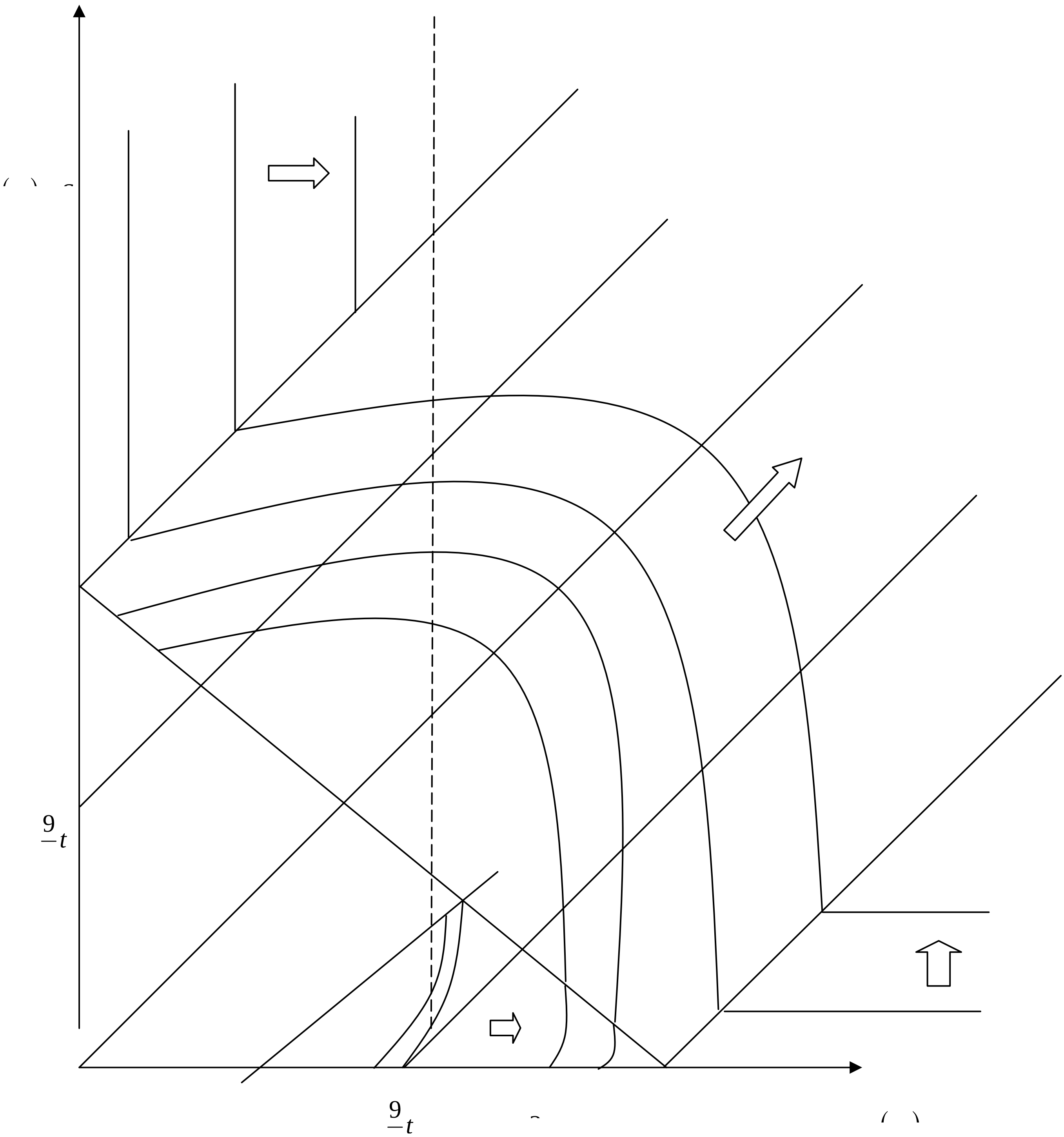
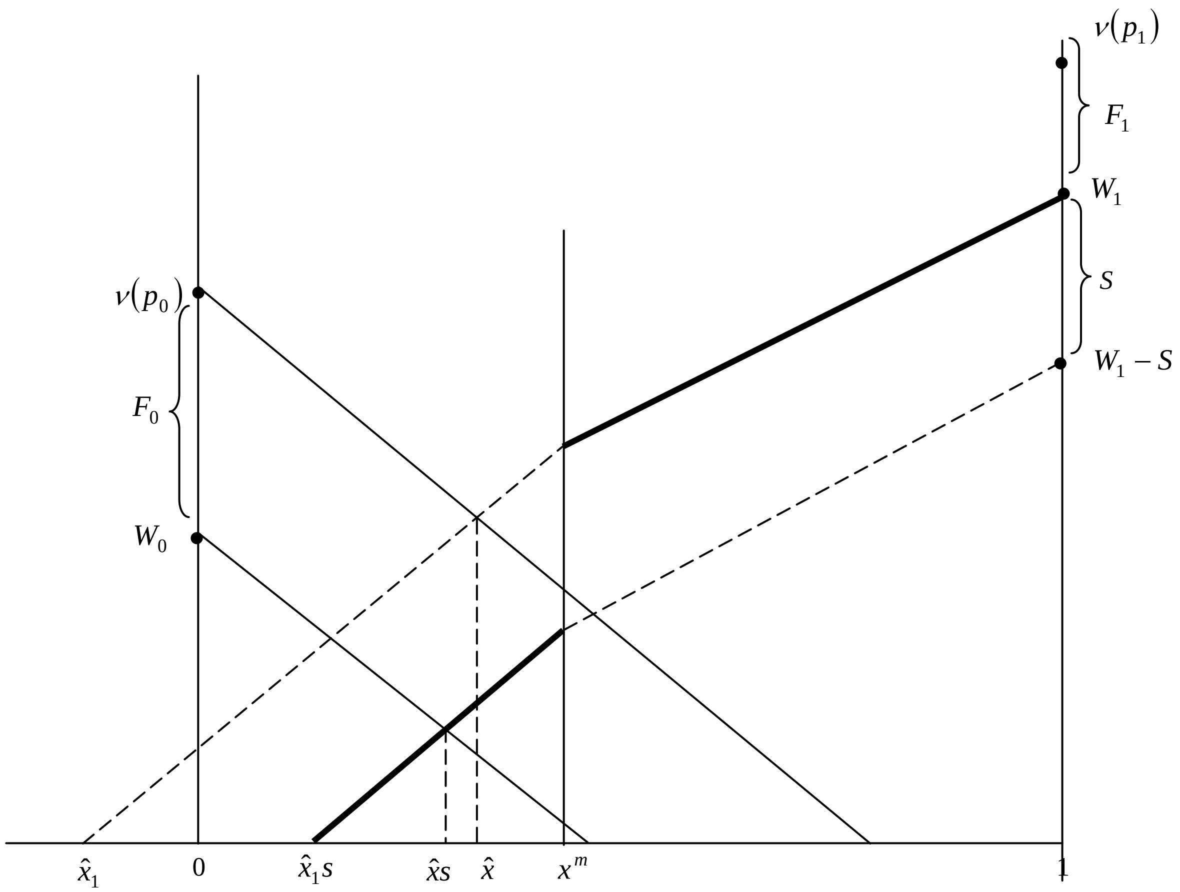


Figure 3: Iso-Social Surplus



Stage 1 **Incumbent sells to** $(0, x^m)$

Stage 2 $x \ \varepsilon (0, x^m)$ **incurs switching cost S to buy from Entrant**

Figure 4: **Infant Network** $\frac{\nu(c_0)}{2} < t$

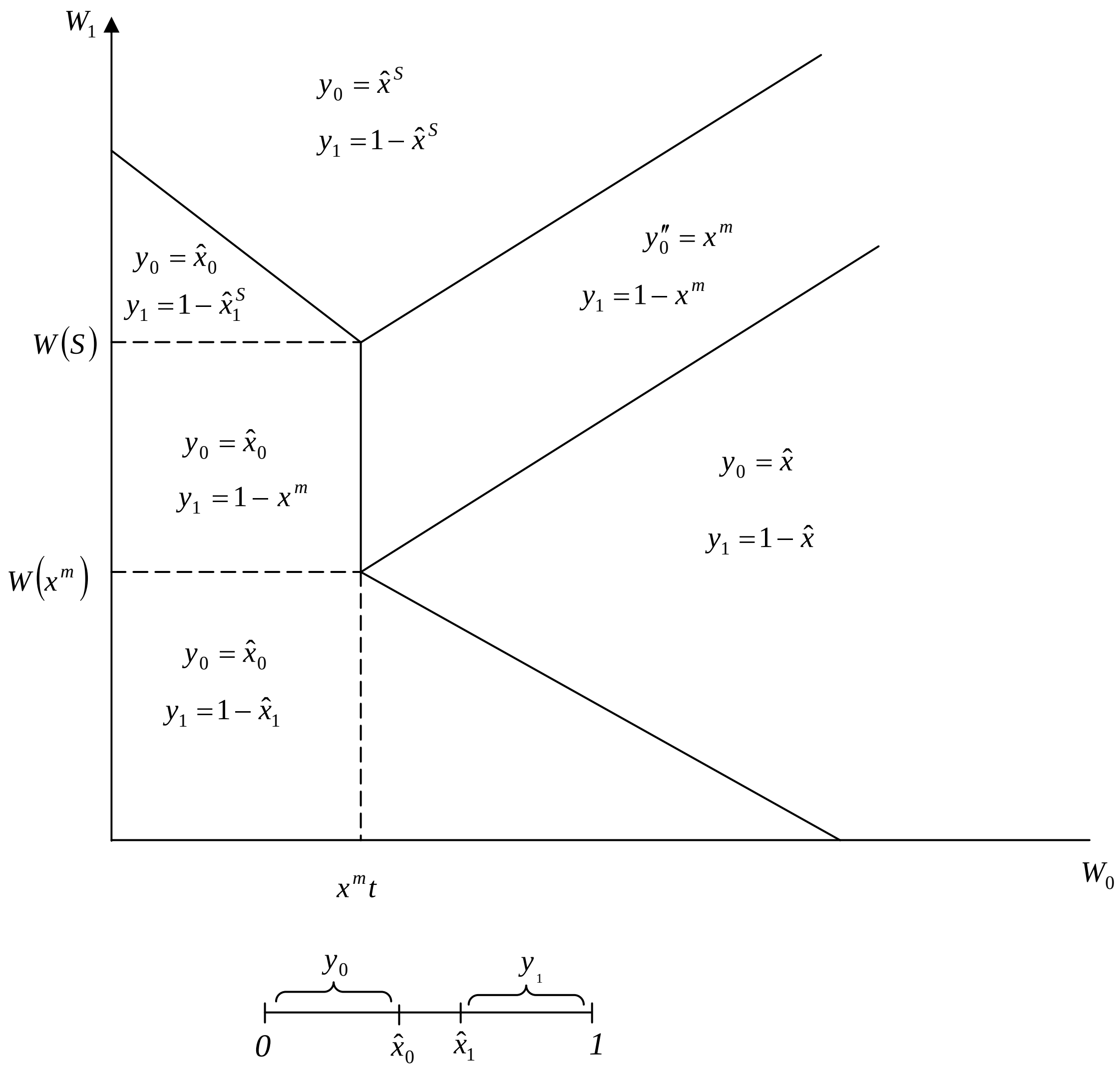


Figure 5: **Demands** y_1 and y_0