Coordination and Lock-In: Competition with Switching Costs and Network Effects

the latest version of this paper, and related material, will be at
www.paulklemperer.org

Joseph Farrell
University of California, Berkeley
549 Evans Hall # 3880
Berkeley, CA 94720-3880
USA

email: farrell@econ.berkeley.edu

and

Paul Klemperer
Nuffield College, Oxford University
Oxford OX1 1NF
England

Int Tel: +44 1865 278588
Int Fax: +44 1865 278557

email: paul.klemperer@economics.ox.ac.uk

First draft: 1997
This draft: December 2001

PRELIMINARY DRAFT: PLEASE SEND COMMENTS

1 Introduction

The economics of switching costs and network effects have received an unusual amount of popular, as well as professional, attention in the last two decades. But while they are thoroughly modern topics with a central place in the "new economics", they are also closely linked to the more traditional concepts of contract incompleteness, complementarity, and economies of scale and scope.

Switching costs arise when a consumer makes investments specific to buying from a firm, creating economies of scope between buying different goods, or (especially) goods at different dates, from that firm. The consumer then values compatibility—the ability to take advantage of the same investment—between his purchases. Network effects arise when a consumer values compatibility—often stemming from ability to take advantage of the same complements—with *other* consumers, creating economies of scope between *different* consumers' purchases.

Both phenomena shift the locus of competition from smaller to larger units of sales, typically extending beyond the sales formally covered by simple sales contracts. While contracts may cover only a single purchase by a single buyer, the buyer really cares about a stream of purchases or about what other buyers are doing: there is a mismatch between contract terms and the outcomes that really determine surplus. When there are switching costs, competition focuses at least partly on streams of products or services, while contracts often cover only the present.

Similarly, when there are network effects, competition largely concerns selling to large groups of users, while contracts usually cover only a bilateral transaction between a seller and one user. When these groups of users do not make simultaneous choices, early choices tend to commit later buyers: there are "collective switching costs." When users do make simultaneous choices, they face a coordination problem. Sufficiently clever contracts can solve these problems, but ordinary contracts may do much less well.

Generally, incomplete contracts with switching costs or network effects cause problems, as the literature explores. However, the problems are more subtle than simply observing that a buyer may be locked in by his own previous choices or by others'. One must account for the competition to acquire such ex post power. In some cases, this implies that contracts are effectively more complete than they seem. For example, in the simplest switching-cost models, initial sales contracts do not specify future prices, yet efficient competition for the stream of purchases can arise. Similarly, in some simple network-effect models, users efficiently coordinate and network effects cause no trouble. In such models, conventional competition "in the market" is replaced by well-functioning competition "for the market"—in the case of switching costs, for a buyer's lifecycle requirements, or in the case of network effects, for the business of enough buyers to exhaust network effects. In both cases early adoptions tend to be pivotal and are thus the focus of competition; later, locked-in buyers may pay more and create the ex post rents for which there is fierce ex ante competition. Aside from possible distributional issues, this may be efficient.

But these models are only the beginning of the story: things do not always work so well. Despite competition for the market, incompatibilities often reduce efficiency and harm consumers in various ways. First, they tie together transactions and thus limit specialization: this may raise costs or reduce the efficiency of buyer-seller matches. Second, consumers may actually switch or actually adopt incompatible products, which has direct costs. Third, although ex post rents created by incompatibility are sometimes fully and efficiently passed through in ex ante competition, they may instead be partly dissipated in inefficient competition, and may be partly captured by an early mover, especially (with network effects) if expectations fail to track relative surplus, and perhaps (with switching costs) if early adoters are myopic. Fourth, the bargains-then-ripoffs price pattern with switching costs, or the analogous pattern of generous offers to pivotal buyers and exploitation of non-pivotal buyers with network effects, normally distorts buyers' quantity choices. Fifth, incentives for entry are often distorted – not necessarily, but often, in the direction of discouraging efficient entry. Finally, firms may dissipate resources creating and defending incompatibility – by contract, product design, or (as in "aftermarket" cases such as Kodak) business practice.

With both switching costs and network effects, *expectations* play a key role. Efficiency and consumer surplus depend on transactions complementary with the current one, but contracts often fail to specify those complementary transactions, so buyers' expectations about them are crucial. At the same time, viewing the links between periods from the other end, *history* matters. At the consumer level, history matters because a buyer may be locked in or may feel a need for compatibility with an installed base; thus past adoptions guide future adoptions. For firms, past *market share* is a valuable asset because (in the case of switching costs) it represents a stock of individually locked-in buyers, or (in the case of network effects) the installed base can boost the network benefits the firm can offer and may also (for that very reason) help attract favorable expectations. Vying for valuable share, firms may compete hard for early adoptions, notably with penetration pricing.

1.1 Switching Costs

A product has switching costs if a buyer will buy it repeatedly and will find it costly to switch from one seller to another during that series. Switching costs also arise if a buyer will make a series of transactions in possibly different products (such as service and repair on an initial purchase) and will find it costly to switch suppliers.

If this lock-in is strong enough, a buyer is effectively buying a series of goods. Not surprisingly, sellers sometimes accordingly offer "life-cycle" contracts. In this case nothing is fundamentally new: with strong enough relationship-specific economies of scope, competition is naturally over bundles of goods rather than over single goods. The difference is that with economies of scope at a single date, contracts are more apt to specify all the prices. With dynamic economies of scope (switching costs), contracts are often short-term. Short-term contracts governing a long-term relationship often create ex post monopoly, for which firms compete ex ante, as in Williamson's (1985) "fundamental transformation."

While we think of switching costs as costs of changing suppliers from one *period* to another, consumers may also face costs of using a different supplier for one good than for *another good* in a single period. These "shopping costs" are a kind of switching costs but are less apt to entail commitment problems. Shopping costs may induce firms to produce a socially-excessive *number* of products, but insufficient *variety.*

Switching costs shift competition away from what we normally think of as the default (a single consumer's needs in a single period) to something broader—a single consumer's needs over time. This need not cause competitive problems: competing on first-period terms can be an adequate proxy for competition on broader terms, if the "bargain-then-ripoff" pattern of pricing low to attract business and then high to extract surplus does not harm efficiency. Likewise, the theory of bilateral contracts with hold-up shows that when parties cannot contract on important future variable and there are switching costs, it can be efficient to accept that hold-up will occur and to compensate the prospective victim up front. Sometimes, even when each firm sets prices to an entire market rather than to a single user, firms engage in repeated "sales" and efficient bargain-then-ripoff pricing survives. But such pricing is often inefficient, potentially causing quantity distortions, wrong incentives to switch suppliers, and incorrect entry incentives, for instance.

In other switching-cost markets, firms set the same price to both old (locked-in) and new customers. Having some locked-in buyers encourages a firm to set relatively high prices; but a countervailing force is that firms might set low prices to build a locked-in customer base. Thus the overall price effects of switching costs in oligopoly are ambiguous; it turns out that (as with network effects) the price effects depend heavily on how consumers form expectations about future prices. But it seems likely that switching costs increase prices, particularly when firms artificially create switching costs to increase their profits.

The effects on entry are not straightforward: switching costs tend to hamper entry if entrants would have to attract existing customers (for instance to achieve minimum viable scale), but may actually encourage small-scale entry that focuses purely on new customers. For much the same reason, stable competition with multiple firms is likely with switching costs: a larger customer base encourages a firm to harvest its base relatively more than to invest in winning new customers; this fat-cat effect lets smaller firms catch up. More generally, the tradeoff between harvesting and investing depends on interest rates, the state of the business cycle, expectations about exchange-rates, etc, with potentially important macroeconomic and international-trade implications.

1.2 Network Effects

A good exhibits network effects if each user's payoff is increasing in the number of other users of that good or of goods compatible with it. From a cooperative game theory point of view, this is nothing new: the surplus available to any coalition of buyers and sellers is the same for the network-effects case as for a corresponding economies-of-scale case.[1] But the coordination issues are much harder.

At the same time, network effects often create an externality. When a single network product competes against non-network alternatives, both coordination problems and the externality (if the network effects are not internalized) push toward underadoption, so (non-deceptive) techniques by which a firm or firms try to affect expectations typically also enhance efficiency. Early literature largely focused on this single-network case.

The modern literature, by contrast, focuses on competition among products *each* of which displays network effects, so any externality is weaker, because adopting one network means failing to adopt another. The primary questions are instead how well competition works in selecting *which* network product dominates and whether such competition allows excess profits or gives consumers the full benefits of competition.

With simultaneous adoption by many buyers, network effects tend to create multiple equilibria: for instance, it might be an equilibrium for nobody to adopt a product, or for everyone to do so. What happens depends on expectations (everyone will do what they expect everyone else to do), and may be unpredictable or, at least, depend on things omitted from a simple model.[2]

With sequential adoption and 20/20 foresight, the outcome may become predictable, but it still has interesting properties. The network effects create positive feedback or "tipping": a product that looks like taking off, or taking most of the market, will *as a result* do so. So competition focuses on the early stages of adoption. The outcome is initially unstable and easily influenced, and perhaps unpredictable, but later becomes locked-in and not readily changed either by random forces or by competition. Thus long-run outcomes may be "non-ergodic" and depend sensitively on seemingly small early chance events.

So firms competing to achieve dominance will heavily court early (or, more generally, pivotal) adopters with good deals; once these adopters have cast their choices, later adopters may be subject to an "ex post monopoly" protected by network effects. The resulting bargains-then-ripoffs price path is all the

[1] Likewise, from a cooperative game theory point of view, a single-consumer model of switching costs corresponds to a model with economies of scope on the production side. But richer switching-cost models do not correspond to simple economies-of-scope cases because the individual customer-supplier matches matter.

[2] Although we focus on buyers' expectations, network effects often operate through complementarities among players at different levels. For example, in computer operating systems, while buyers probably do prefer an operating system that attracts more other buyers as such, they also like an operating system that attracts applications developers — which it will tend to do if it attracts many buyers. The recent Microsoft litigation stresses this indirect network effect, arguing that it creates an applications barrier to entry if users' and applications developers' expectations focus on Microsoft's product without strict comparisons between its price and intrinsic quality and those of other operating systems. Another example is in credit cards: a network that signs up many cardholders finds it easier to recruit merchants, and vice versa.

more dramatic if later adopters' expectations will track past success (installed base) rather than optimally coordinate given the installed base. Because expectations and adoptions are self-reinforcing, often only one network can survive in incompatible competition; weak horizontal differentiation may as well not exist at all, as adopters all focus on joining the expected winner.[3]

All this dramatically affects the look and feel of how markets behave. If buyers were identical and could smoothly coordinate, none of it might matter for efficiency and consumer surplus—just as with identical buyers and without network effects, the seller who offers the most surplus makes all the sales, and competition to do so ensures buyers a competitive level of surplus.[4] But if coordination is hard or buyers' tastes differ, they may coordinate only clumsily, or not at all. Clumsy coordination, on which most of the literature focuses, may put only weak competitive pressure on sellers to offer good deals, if each seller perceives its chance of winning the market as depending only mildly on the surplus it offers. All this depends sensitively on how buyers coordinate: for instance, competition works very differently if buyers' expectations of what others will do track surplus than if expectations stubbornly favor one firm.

Institutions, cheap talk, and sequential (rather than simultaneous) choice can all help focus expectations to the benefit of adopters; but a seller of a network product can also affect buyers' expectations, for instance by penetration pricing, marketing, or strategic choices such as second-sourcing.

So, as with switching costs, competition for the market *can* be efficient and vigorous—but incompatible network effects will often induce inefficiency and monopoly. If later adopters want to be compatible with earlier adopters, and the latter face switching costs, there are "collective switching costs" for the market as a whole. So the dynamics and problems of switching-cost markets are largely inherited here as well, with the added problem that early adopters are not the same people who will later be locked in. But competition between incompatible networks can also be obstructed if adopters fail to resolve their coordination problems, an issue that we frame in terms of expectations. For instance, if expectations systematically favor incumbents, entry will tend to be (as it can be anyway) even harder than it should be given incompatibility and given the installed base.

Firms collectively (and sometimes policy-makers) choose whether they compete with compatible or incompatible products. Efficiency, relative advantage, and competition-limiting motives can all be part of their calculations: in particular, firms who expect to attract expectations (other than by making a better offer) tend to favor incompatibility to weaken the perceived value to users of rivals' offerings.

1.3 Why is this interesting?

These topics have attracted a lot of public attention for scholarship in industrial organization (they have featured in *Scientific American* and even the *New Yorker*). What accounts for the popular appeal? We think it is a combination of scientific, social science, business strategy, and policy appeals.

At a general scientific level, notions of positive feedback and unstable dynamics have attracted a great deal of recent attention. Physical scientists, perhaps partly freed from analytically tractable (typically stable) models by improved computing power, are increasingly interested in complex dynamic systems; the Santa Fe Institute has been a visible focus of such research. Lock-in through switching costs and network effects may make outcomes depend even in the long run on small events and accidental happenstance, while positive feedback can make long-run outcomes paradoxically both rigid and unpredictable. While the advanced mathematics of positive feedback and chaos theory has played little part in the economic literature we address, the Zeitgeist may contribute to the literature's appeal; it is no coincidence that Brian Arthur, the first economist to direct the Santa Fe Institute, was one who had worked on network effects.

[3] And if network effects grow stronger over time, previously lucrative niche products may cease to be viable.

[4] Of course, one must not confuse the winner-take-all outcome with monopoly power, but this point applies equally to the case of identical buyers and no network effects but ordinary scale economies.

Our topics are interesting to social scientists in part because they assign an important role to the details of social institutions. When a standards organization makes a choice, or early buyers pick a network product, or a procurement officer makes an initial purchase, they and others may be committed for considerable periods. Decisions are less autarkic and less reversible than in a conventional convex economy. This puts greater weight on the institutions that make initial or collective choices, and on how they make the choices. And as we noted above, history also matters more dramatically and in a more interesting way than in conventional economic theory.

For students of business strategy, network effects and lock-in raise richer competitive strategy issues than those arising in a vanilla industry. They are particularly important in some of today's most interesting and widely reported industries — the information industries — and it is often argued that the new economy revolves more around these issues than did an older economy.[5] Moreover, the issues interact closely with those of intellectual property, because intellectual property rights can often help a firm unilaterally to impose incompatibility and switching costs. In addition, the economics of network effects and switching costs illuminates how firms' market shares matter, helps explain business strategies such as penetration pricing, and resonates with the intuitive notion that firms build and then exploit share.

Finally, the topics addressed here are live ones in antitrust and competition policy. Switching costs have been important in a range of antitrust cases, including IBM and Kodak. The network effect by which popular operating systems attract applications software supply was central in the Microsoft case. Both network effects and switching costs shift the locus of competition from textbook competition in the market to some form of Schumpeterian competition for the market, in which firms compete to achieve dominance. The relationships between these forms of competition, and the extent to which society should be happy with such a shift in the locus of competition, have not been fully explored, which leaves room for a great deal of professional disagreement over policies for such markets. We hope our discussion may help focus, and perhaps even defuse, such disagreement.

2 Switching Costs and Competition

2.1 Introduction

A consumer faces a *switching cost* between sellers when an investment specific to his current seller must be duplicated for a new seller. That is, the switching cost is caused by the consumer's desire for compatibility between his current purchase and a previous investment. That investment might be in equipment, in setting up a relationship, in learning how to use a product,[6] in buying a high-priced first unit that then allows one to buy subsequent units more cheaply, or even psychological.[7]

[5] A note of caution is worthwhile: to the extent that the issues are more live before a dominant design emerges, we are necessarily more likely to see standards battles in fast-growing new-technology industries than in settled industries at any time, not just now.

[6] Similar effects to switching costs arise when consumers face uncertainty about the characteristics or qualities of products that they have not tested (see for example, Schmalensee (1982) and Villas-Boas (2000)), or when firms face uncertainty about consumers they have not supplied and can exploit those they know to be of "high quality" (see for example, Nilssen (2000) on insurance markets and Sharpe (1990) and Zephirin (1993) on bank loan markets).

[7] Although there is a large theoretical literature on switching costs, much less empirical work has been done in this area, perhaps because switching costs are often consumer specific and not directly observable, and yet must often be estimated from aggregate data because micro data on individual consumers' switching are not commonly available.

The empirical literature includes Ausubel (1991) on the credit-card market, Borenstein (1991) on retail gasoline, Beuhan (1997) on the windows and DOS operating systems for personal computers, Dahlby and West (1986) on liability insurance, Elzinga and Mills (1990) on the wholesale distribution of cigarettes, Greenstein (1993) on the procurement by federal agencies of mainframe computer systems during the 1970's, Kim, Kliger and Valoe (2000) on the market for bank loans, Knittel (1997) on the market for long-distance phone-calls, Moshkin and Shachar on television viewing choices (1999), Schlesinger and Schulenburg (1993) on automobile insurance, Sharpe (1997) on the bank deposit market, Shum (1999) on cereals, Shy (2000) on the cellular-phone market and on the market for bank deposits and Stango (1998) on credit-cards.

[TIDY UP] Experimental studies are even fewer, but include Cason and Friedman (2000).

Switching costs can be *informational,* as in the case of learning costs. Then a consumer who switches from firm A to firm B has no switching cost of later buying from either firm. Alternatively, switching costs can be *transactional*, such as the cost of returning rented equipment and renting from a new supplier. Then a consumer who switches from A to B would incur an additional switching cost if he reswitched back to A.[8]

We will generally assume that switching costs are real social costs, but there can also be *contractual* or purely pecuniary switching costs (that are not social costs), such as airlines' "frequent-flyer" programs, and "loyalty contracts" that, for example, rebate a fraction of past payments to consumers who continue to patronise the firm.[9]

Switching costs do not only apply to repeat-purchases of identical goods. In many important examples "follow on" goods, such as spare parts and repair services, are bought in "aftermarkets," and there are "switching" costs if the follow-on goods are not compatible with the original purchase, as may be the case if they are not bought from the same firm.[10] [11]

Consumers may also incur switching costs, or "shopping costs", at a single date by buying related products from multiple suppliers rather than from a single supplier. However, most of the literature focuses on dynamic problems and emphasises the resulting commitment problems.

In the simplest cases, when firms can commit to future prices and qualities, etc., a market with switching costs is closely analogous to a market with economies of scope in production; with switching costs each individual consumer can be viewed as a "market" with economies of scope between "purchases now" and "purchases later". Just as a market with production economies of scope is entirely captured by the firm with the lowest total costs in the simplest price-competition model, so in a simple model with complete contracts each individual buyer's entire stream of needs in a market with switching costs is captured by the lowest-cost supplier of his whole lifetime's requirements. That is, firms compete on their "lifecycle" prices[12] and the market price is determined by the second-lowest cost firm, and the outcome is efficient and switching costs confer no market power on firms.

But this simple analogy—and very often the efficiency of the outcome—breaks down if firms cannot credibly commit to future prices or qualities (Section 2.2). The analogy is still weaker if firms cannot discriminate between different customers (Section 2.3), or consumers use multiple suppliers (Section 2.4). We treat these cases taking both the switching costs and the number of firms as exogenous, before considering new entry (Section 2.5), endogenising switching costs (Section 2.6), analysing the "market share" competition that switching costs can generate (Section 2.7), and finally addressing the implications for competition policy (Section 2.8).

[8] Nilssen (1992) first emphasised this distinction. Of course, many switching costs fall between the extremes. Moshkin and Shacher (1999) use panel data on television viewing choices to estimate the proportions of consumers who behave as if they are fully informed but face transactional switching costs, and the proportion who behave as if they face informational switching costs.

[9] Klemperer (1995) gives many other examples of each of these kinds of switching costs.

[10] *Aftermarkets* have been much studied since a US Supreme Court decision (ITS v. Kodak, [cite]) held that it was conceptually possible for ITS, an independent repair firm, to prove that Kodak had illegally monopolized the aftermarket for servicing Kodak photocopiers, see e.g. Shapiro and Teece (19xx), MacKie-Mason (19xx), and Borenstein, MacKie-Mason, and Netz (2000). [further work to cite.]

[11] Typically, a consumer who has not previously bought from any firm incurs a start-up cost similar to the new investment (switching cost) that a brand switcher must make. We will use the term "switching cost" to include these start-up costs. So a consumer may have a "switching cost" of making a first purchase; in many models this is of little importance since it has no effect on consumers' preferences between firms.

Many of consumers' costs of switching to new suppliers have parallels in firms' costs of serving new customers (see Klemperer (1995)). These have been less studied, but in some contexts, such as the simple model of the next subsection, the total prices (including any switching costs) paid by consumers, and the implications of switching costs are unaffected by whether firms or consumers actually pay the switching costs.

[12] Indeed consumers (and salesmen) often do focus on "lifecycle" prices when considering capital purchases such as photocopiers.

2.2 Firms who Cannot Commit to Future Prices

2.2.1 Bargains Followed by Ripoffs

The core model of the switching costs literature posits that firms cannot commit to future prices.

The simplest model has two periods and two firms, with costs c_i in periods $i = 1, 2$.[13] A single consumer has a switching cost s and reservation price $r_i > c_i + s$ for one unit of the period$-i$ good, firms set prices, and there is no discounting. Then in period 2 the firm that sold in period 1 will exercise its ex post market power by pricing (just below) $c_2 + s$ (the rival firm will offer price c_2 but make no sale). Foreseeing this, firms are willing to price below cost in period 1 to sell to the customer who will become a valuable follow-on purchaser in period 2; competition to win the customer will drive period-1 prices down to $c_1 - s$.[14] [15]

Although first-period prices are below cost, there is nothing predatory about these prices in this context, and this pattern of low "introductory offers", or "penetration pricing" (see section 2.7), followed by higher prices to exploit locked-in customers is a familiar one. Klemperer (1995) gives many examples. "Bargains-then-ripoffs" is a main theme of the switching-costs literature.[16]

Although the switching cost strikingly affects price in each period, it does not affect the life-cycle price $c_1 + c_2$ that the consumer pays. As in section 2.1, we can here view the life-cycle (the bundle consisting of the period-1 good and the period-2 good) as the real locus of competition, and competition in *that* product has worked exactly as one would hope. In particular, the absence of price commitment did not lead to any inefficiency in this very simple model.

2.2.2 Inefficiency of the Price-Path

Although the outcome above is socially efficient, the inability to contract in period 1 on period-2 prices in general leads to inefficiencies. In particular, if we modify the model above so that the consumer has downward-sloping demand in each period and firms are restricted to linear pricing (i.e. no two-part pricing) the result would be socially excessive consumption in period 1 and too little consumption in period 2.[17] More generally, if there are many heterogeneous consumers, and firms cannot distinguish between them, there will be excessive sales in period 1, and too few sales in period 2.[18]

Our simple model also assumed firms can in the introductory period give rents to the consumers whom they will exploit later. In practice this may not be possible. For example, setting very low introductory

[13] $c_2 \neq c_1$ is especially natural if the second-period good is spare parts/repair services/consumables for a first-period capital good.

[14] This argument suggests that firms should charge lower prices to their rivals' customers than to their own customers. See also, for example, Chen (1997), Taylor (1998), Lee (1997): Shaffer and Zhang (2000) show that if firms are asymmetric, one firm may charge lower prices to its own customers.

[15] [Note 2.2.1]. Because firms are symmetric and so charge the same price in period 2, the consumer is clearly indifferent in period 1. If firms A, B had different costs c_{A2} and c_{B2} in period 2, then if A made the period-1 sale, its period-2 price would be $p_{A2} = c_{B2} + s$ (that is, constrained by B), while if B made the period-2 sale, its period-2 price would be $p_{B2} = c_{A2} + s$. In this case, the prices that firms charge in period 1 (and hence also firms' incentives to invest in cost reduction, etc.) depend on whether the consumer has rational expectations about the period-2 prices it will face or whether the consumer acts myopically. We discuss the role of expectations in Section 2.3.5. Other simple models such as that in Klemperer (1995, Section 3.2) sidestep the issue of consumers' expectations by assuming period-2 prices are constrained by consumers' reservation price r_2, hence independent of consumers' period-1 choice. The distinction between these modelling approaches is crucial in some analyses of network effects (see Section 3.7.3).

[16] Two-period models that make this basic point include Klemperer (1987a, b), Basu and Bell (1991), Padilla (1992), Basu (1993), Ahtalia (1998), and Pereira (2000). Skott and Jepsen (2000) argue that a tough drug policy may encourage the aggressive marketing of illegal drugs to new users, by increasing the costs of switching between dealers.

[17] Thus discussions of aftermarket power point out the possibility of wrong tradeoffs between aftermarket maintenance services, self-supplied repair, and replacement of machines. See Borenstein, MacKie-Mason, and Netz (199x), for instance.

[18] See Klemperer (1987a). Nilssen's (1992) model distinguishes between informational and transactional switching costs, and shows that the lower incentive for consumers to switch in the transactional case can lead to lower prices for new consumers and higher prices for loyal consumers and so lower social welfare than in the informational case, when firms can distinguish the histories of different consumers.

prices may attract worthless customers who will not buy after the introductory period.[19] If for this or other reasons firms dissipate their future profits in unproductive activities (e.g., excessive advertising and marketing), consumers are made worse off by switching costs even if competition ensures that firms continue to make zero profits over the two periods.

In our simple model firms make zero profits with or without switching costs. In more general two-period models oligopolists' profits can be either raised or lowered by the higher ex-post prices and lower ex-ante prices that switching costs create—see Klemperer's (1987a, b) models covering both price and quantity competition and many consumers.

Finally note that while our discussion (and the literature's) is in terms of exploiting locked-in consumers with high prices, they can equally be exploited with low qualities; indeed future quality is even harder to contract on than future price, and the inability to contract on quality may make price commitments pointless.[20]

2.3 Firms who Cannot Discriminate Between Cohorts of Consumers

In our first bargains-then-ripoffs model, we assumed firms could distinguish between "old" and "new" consumers, perhaps because "old" consumers are buying "follow on" goods such as spare parts. But when old consumers buy the same good as new consumers, it is often difficult for firms to distinguish between them.

2.3.1 Free Entry Model

Even if firms cannot distinguish between cohorts of consumers, we may get the same pricing pattern if firms specialize sufficiently. In particular, with free entry of identical firms, in each period some firm(s) will specialise in selling to new consumers while any firm with any old locked-in customers will sell only to those old customers.

If consumers have probability ϕ of surviving into each subsequent period new-entrant firms with constant marginal costs c and discount factor δ offer price $c - \phi\delta s$ and sell to any new consumers, while established firms charge s more, that is, charge $c + (1 - \phi\delta)s$ in every period.[21] No "old" consumers switch, and new entrants' expected discounted profits are zero. Thus the price paths consumers face are exactly as if firms could perfectly discriminate between them. In either case one can think of every (new and old) consumer as getting a "discount" of $\phi\delta s$ in every period reflecting the present value of the full extent to which he can be exploited in the future by his existing seller, given his option of paying s to switch to an entrant; simultaneously, every "old" consumer is indeed exploited by s in every period. The outcome is socially efficient.

2.3.2 Price Patterns in a Closed Oligopoly

Just as in the free-entry model, if there is a small number of firms who face no threat of entry and who cannot distinguish between cohorts of consumers, it is possible that in every period one firm might hold a "sale", setting a low price to attract new consumers, while the other(s) set a higher price to exploit their old consumers. Farrell and Shapiro (1988) explore such an equilibrium, but their model in effect has just

[19] There may also be limits on firms' ability to price discriminate in favour of new customers without for example dangerously antagonizing their "regular" customers. See Section 2.3 for the case in which price-discrimination is infeasible.

[20] Farrell and Shapiro (1989) show that price commitments may actually be worse than pointless. See note FS [Section 2.6].

[21] This assumes all consumers have reservation prices exceeding $c + (1 - \phi\delta)s$ for a single unit in each period, and that all consumers survive into the next period with the same probability, ϕ, so a consumer's value is independent of his age. If consumers live for exactly two periods the price paths in general depend on whether firms can directly distinguish between old and new consumers (as in the previous subsection) or cannot do this (as in this section).

one new and one old consumer in each period. Since it is therefore inevitable that at the beginning of any period one firm has no customer base while the other already has half the market "locked-in", it is not surprising that this model yields firms offering asynchronous sales.

More generally it seems unclear whether oligopolists will hold sales simultaneously or not. For example, one could argue that it makes more sense to forgo short run profits to go after new customers when your rivals are not doing so. But one might alternatively argue that if switching costs are informational, then staggered sales cause switching between firms and thus create a pool of consumers who have no further costs of switching, thus intensifying future competition.[22] Indeed Klemperer (1983, 1989) and the extension of the latter model in Elzinga and Mills (1999) all have simultaneous sales.[23] [24]

Another possibility is that rather than holding occasional sales, each oligopolist in every period sets a single intermediate price that trades off its incentives to set a low price to attract new consumers and to set a high price to exploit its old customers. In a steady-state model each firm's price will then be the same in every period. This can only be an equilibrium under certain conditions: If the flow of new consumers is too large, at least one firm would deviate from any candidate equilibrium by cutting price significantly to specialise in new consumers. If some consumers' switching costs and reservation prices are too large, at least one firm would deviate by raising price significantly to exploit old customers while giving up on new ones. And if firms' products are undifferentiated except by switching costs, there cannot easily be a symmetric pure-strategy equilibrium because each firm would have incentive to undercut the other by a tiny amount to win the new consumers.[25] But when none of these conditions applies, there may be a stationary "no-sales" equilibrium, and much of the literature examines such equilibria. [26]

2.3.3 Industry Dynamics

We have seen that sometimes a small firm holds a sale while its larger rivals do not. Similarly, in most no-sale models in which all firms sell to both old and new consumers, firms with more old locked-in customers have a greater incentive to exploit them and therefore price higher and so win a smaller share of new unattached consumers. In both cases, the result is stable industry dynamics. In Beggs' and Klemperer's (1992) no-sale model, for example, the duopoly converges monotonically back to a stable steady state after any shock.

---

[22] See section 2.4. Most models make assumptions that eliminate this effect, although Klemperer (1992) emphasises it.

[23] Elzinga and Mills' model fits with observed behaviour in the cigarette market.

[24] Padilla (1995) gives an ambiguous answer to whether firms synchronise or stagger their sales by yielding mixed strategy equilibria. [But I think this mixing is driven by neither of these "effects"?].

Numerous papers (Baye et al (1992), Padilla (1992), Deneckere et al (1992), Fisher and Wilson (1995), Green and Scotchmer (1986), Rosenthal (1980), Shilony (1977), Varian (1980), analyse single-period models of switching costs (or models that can be interpreted as being of switching costs) that yield mixed strategy equilibria. However, adding more real-world features to some of these models yields either asymmetric pure-strategy equilibria or symmetric pure-strategy Bayesian-Nash equilibria (if information is imperfect) rather than mixed-strategy equilibria.

The former (assymetric) outcome can be interpreted as asynchronous sales. Deneckere et al obtain this kind of equilibium when they allow firms to choose the times at which they set their prices; they find that the firm with fewer locked-in customers sets price second and holds a "sale".

The latter (symmetric) outcome corresponds to "tradeoff pricing" of the kind discussed in the next paragraph. Bulow and Klemperer (1998, Appendix B) give an example of this by incorporating incomplete information about firms' costs into a one-period model of switching costs that would otherwise yield mixed-strategy equilibria.

[25] However, if consumers have rational expectations about future prices, a small price cut may win only a fraction of new consumers. We discuss consumers' expectations below.

[26] The leading "no-sales" models are von Weizsäcker, 1984, Beggs and Klemperer, 1992, the variants of the latter model in Chow, 1995, and To, 1995 and Wernerfelt (1991); Phelps and Winter's (1970) and Sutton's (1980) models of search costs are related.

Even if there are occasional "sales" the need to balance exploiting the old with attracting the new will be present in "ordinary" periods, and this literature is relevant to these ordinary periods.

In the case of monopoly, both stationary "no-sales" models (see Holmes, 1990) and models with periodic sales (see Gallini and Karp, 1989) are possible.

However, the opposite is possible. If larger firms have lower marginal costs, and especially if economies of scale make it possible to drive smaller firms completely out of the market, then a larger firm may charge a lower price than its smaller rivals. In this case, any small advantage one firm obtains can be magnified and the positive-feedback dynamics can result in complete dominance by that firm. This is just as is typical with network effects (see Section 3.X)—indeed, switching costs create positive network effects in this case, because it is more attractive to buy from a firm that other consumers buy from (Beggs, 1989).

So switching-costs markets *can* "tip" like network-effects markets. But the simple models suggest a presumption that markets with switching costs are stable, with larger firms acting as less-aggressive "fat cats".[27]

2.3.4 The Level of Prices

Is the symmetric stationary price of a "no-sales" equilibrium higher or lower than if there were no switching costs?

If products are otherwise undifferentiated and firms compete in prices, then they earn no profits without switching costs, while Padilla (1995) shows that they can earn positive profits with switching costs. However, the answer is less obvious if firms' products are differentiated in their "real" (functional) characteristics as well as by switching costs, so that firms would have market power even without switching costs. We therefore focus on this latter case.[28]

Price levels may be *raised* by switching costs through the "fat cat" effect discussed above: firms price less aggressively because they recognise that if they win fewer customers today, their rivals will be bigger and (in simple models) less aggressive tomorrow.

How consumers form expectations also critically affects the price-level: as we now discuss, different assumptions about expectations imply very different conclusions, but the general presumption is that switching costs raise prices.

2.3.5 Expectations

Switching costs make contracts less complete than is efficient: a consumer cares about a seller's future prices in making his purchase decision today, but he and the seller can't agree those prices today. So, in comparing current offers, a consumer is heavily influenced by what he thinks those offers imply about future prices. Thus how consumers use current prices (including any departures from the current prices that they had expected to see) in forming expectations about future prices critically affects competition and the price level—just as in other parts of the lock-in literature.[29]

Consumers' expectations about their own future tastes also matter. We assume consumers expect that their tastes may perhaps change, but that there is some positive correlation between current and future tastes.

(i) *Consumers who Assume any Price Cut below their Expected Price will be Maintained in the Future*

If consumers expect that a firm that cuts price today will maintain that price cut in the future, switching costs can *lower* equilibrium prices. The reason is that consumers recognise that the product they choose

[27] In the terminology introduced by Fudenberg and Tirole (1984). In the terminology introduced by Bulow, Geanakoplos, and Klemperer (1985), there is strategic complementarity between a firm's current price and its competitors' future prices. See also Farrell (1986).

[28] See, especially, von Weizsäcker (1984) and Beggs and Klemperer (1992). Chow (1995) and To (1995) analyse variants of Beggs' and Klemperer's model.

[29] Consumers' expectations about how future prices depend on costs are, of course, also important in determining whether firms have the correct incentives to invest in future cost reduction. This issue does not seem to have been directly addressed by the switching-costs literature, but we discuss in Section 3.7 how a network-effects model can be reinterpreted to address it. See also [note 2.2.1].

today will, because of the switching costs, probably also be the product they choose tomorrow. So relative to the case of no switching costs, when any choice they make is just for today, they are more influenced by a price cut which they expect to be forever, than by their current product preferences. Put differently, in the limit with infinite switching costs, consumers are choosing a product for the whole of time and, if they do not know their future preferences, products are in effect much less differentiated.

For this reason, von Weizsäcker (1984) found that switching costs reduced prices in his model in which—perhaps artificially—firms commit to future price (and quality) paths even though they cannot discriminate among consumers.[30][31]

We will see that a similar effect arises when there are strong network effects and differentiated products (see Section 3.X). With network effects and incompatible competition, consumers' desire to be compatible with others overwhelms their differences in tastes and, when expectations track surplus, drives the firms towards Bertrand competition. Here, with switching costs, each consumer's desire to be compatible with his future self (who in expectation has tastes closer to the average) likewise drives the firms towards undifferentiated-products Bertrand competition.

(ii) *Consumers whose Expectations about Future Prices are Unaffected by Current Prices*

If consumers expect that a firm that sets an unexpectedly low price this period will return to setting the expected price next period, then switching costs *raise* price levels. Each consumer is making a product choice that his future selves must live with, and his future selves' preferences (while possibly different from his own) are likely to be closer to his currently-preferred product than to the other product. So consumers are less influenced by a current price cut than absent switching costs.

(iii) *Consumers with Rational Expectations*

If consumers have fully rational expectations they will recognise that a lower price today generally presages a higher price tomorrow. As we discussed above, a firm that wins more new consumers today will be a "fatter cat" with relatively greater incentive to price high tomorrow.[32] So consumers with rational expectations will be insensitive to price cutting, and prices are *raised* by switching costs.[33]

In summary, while there is no unambiguous conclusion, Beggs and Klemperer (1992) conclude that there is a presumption that switching costs raise prices when new and old customers are charged a common price[34], and that switching costs therefore also generally raise oligopolists' profits.

[30] The effect we discussed above—that firms moderate price competition in order to fatten and so soften their opponents is also eliminated by von Weizsäcker's precommitment assumption.

[31] A related model with these consumer expectations is Borenstein, Mackie-Mason and Netz (2000) in which infinitely lived consumers purchase a differentiated durable good ("equipment") in one period from one of two firms and an aftermarket product ("service") in the next period which must be purchased from the firm from which the durable was purchased. High service prices generate profits from locked-in customers purchasing service, but deter new customers from purchasing equipment because the latter expect high service prices in the following period. So the stationary equilibrium service price lies between marginal cost and the monopoly price.

[32] This is true in existing models, e.g., Beggs and Klemperer (1992), Klemperer (1987a,b,c), Padilla (1992, 1995). One caveat is that, as discussed above, the fat cat effect can sometimes be reversed. Another caveat is that with incomplete information about firms' costs a lower price might signal lower costs, so consumers might rationally infer that a lower price today presaged a lower price tomorrow. On the other hand, incomplete information about costs could also result in firms signalling to each other and, since firms might signal higher costs to their rivals in order to soften future competition, this might be a force for higher prices.

[33] Holmes (1990) analyses price-setting by a monopolist facing overlapping generations of consumers who must sink costs before using the monopolist's good. He finds that if consumers have rational expectations, then prices are higher than those that would prevail if the firm could commit to future prices for essentially this reason: rational consumers are insensitive to price cuts because they understand that a low price today will encourage consumers to sink more costs which in turn results in higher prices in subsequent periods.

[34] See Borenstein (1991) and Ausubel (1991) for empirical evidence supporting this presumption in the markets for gasoline

2.3.6 Collusive Behavior

The arguments above do not consider "collusive behavior" in which high prices are supported by firms punishing any other firm thought to have deviated.[35]

With complete information, switching costs make deviating from a collusive agreement less profitable because it is harder to quickly "steal" another firm's customers, but for the same reason switching costs make it harder to punish a deviating firm. So it is not obvious whether collusion is easier or more difficult than absent switching costs, and in Padilla's (1995) model, switching costs actually make collusion harder.

With incomplete information, however, switching costs may possibly make it easier for firms to monitor collusion, because only large price changes can win away many of a rival's customers, and such large changes may be very easy to observe. Switching costs may perhaps also facilitate collusion by breaking up a market into well-defined submarkets of customers who bought from different firms thus providing "focal points" for tacitly collusive division of the market.[36] But these arguments are weaker if there are many new consumers who are indistinguishable from old locked-in consumers. And if collusion is only easier after most customers are already locked-in, this will induce fiercer competition prior to lock-in. While many people's intuition seems to be that switching costs should support collusion, this intuition does not seem easy to formalise.

2.4 Consumers Who Use Multiple Suppliers

In the models above, as in most leading models of switching costs, switching costs affect prices but there is no switching in equilibrium.[37] In reality a consumer may use more than one supplier, either because the differences between firms' prices to the consumer change over time, or because firms' products are differentiated and the consumer has changing tastes. Furthermore, although we have thus far assumed each consumer buys at most a single unit from a single firm in each period, a consumer who values variety may buy multiple products even in a single period. Consumers may therefore use multiple suppliers in a period or, as we will discuss, firms may each produce multiple varieties. We first ask whether there is socially too much or too little switching, taking the switching costs as given.

2.4.1 Is There Too Much Switching?

Consumers decide whether or when to switch, and pay the switching costs. For them to make socially correct switching decisions, the switching costs they bear must be real social costs and the differences between firms' prices must equal the differences between firms' costs and be unaffected by consumers' decisions.[38] In particular, there will be the wrong amount of switching if (i) consumers switch (or not) in order to affect firms' future prices or (ii) firms' relative prices to a consumer fail to reflect their relative marginal costs or (iii) the switching costs are not real social costs.

(i) *consumers switch in order to affect prices*

If a consumer is a large fraction of the market, or if firms can discriminate between consumers (so each consumer is, in effect, a separate market), a consumer may switch to affect future prices.

---

and credit cards, respectively.

[35] For example, Beggs and Klemperer assume each firm's price depends only on its current market share and not otherwise on history, thus ruling out the kind of strategies described by, for example, Abreu (1988) or Green and Porter (1984) that support collusive outcomes in contexts without switching costs.

[36] These arguments are discussed in Klemperer (1987a).

[37] Indeed, these models starkly make the point that seeing no actual switching between two firms does not mean they do not compete, contrary to simplistic arguments that have sometimes been made in policy debates.

[38] Consumers must also have rational expectations about future price differences, etc.

If switching costs are informational, switching strengthens a consumer's outside option, so if each consumer is a separate market we might observe excessive switching as consumers switch in order to strengthen their bargaining positions. On the other hand, if firms cannot discriminate between consumers, there is an opposite effect: with informational switching costs a consumer who switches usually lowers prices and so improves the efficiency of other consumers' trades with sellers, so there may be less switching than is socially desirable.

Taylor (1998) shows that if switching costs are transactional, consumers may move between suppliers to credibly signal their (already) low switching costs and so secure more favourable terms of trade. Because this switching is socially costly, Taylor shows that equilibrium contracts may discourage it through "loyal customer" pricing policies that give better terms to loyal customers (who have always purchased from the same firm) than to other customers.[39]

Consumers may commit to ignore switching costs in making future purchase decisions, in order to force the incumbent supplier to price more competitively (Cabral and Greenstein (1990));[40] this will generally lead to socially excessive switching *ex post*.

(ii) *price differences don't reflect cost differences*

In our basic "bargains-then-ripoffs" model of section 2.2, firms fully exploit their locked-in consumers in the second period. But if firms do not know their customers' exact switching costs, or if firms do know the switching costs but must set a single price to consumers with different switching costs, they will price too high to some of their old customers and so induce inefficient switching. See, e.g., Chen (1997), Lee (1997), Taylor (1998), Arbatskaya (2000), Gabrielson and Vagstad (2000), Shaffer and Zhang (2000).[41]

Put differently, when firms exploit ex-post market power, their prices and—importantly—the differences between their prices to a given consumer fail to reflect their costs, so inefficient switching can result. And this arises even when firms price symmetrically if firms price discriminate between old and new consumers, because prices to any given consumer are then *not* symmetric.

Furthermore, as we saw in Section 2.3, when firms do not price discriminate between new and old consumers, firms with larger customer bases will charge larger markups over their marginal costs. So if consumers have differing switching costs, the incumbent or dominant firm charges a price that exploits its old high switching-cost customers while allowing its low switching-cost consumers to switch to a smaller firm or new entrant (Gabszewicz, Pepall and Thisse (1992), Farrell and Shapiro (1988), Wang and Wen (1998)).

In all these cases there is excessive switching to smaller firms and entrants.

(iii) *switching costs are not real social costs*

If switching costs are contractual, and not social costs, consumers will presumably switch less than is efficient.

---

[39] But although loyal customers are favoured over old customers who patronised other firms in the past, Taylor also finds firms discriminate in favour of new customers and against old customers for the standard reason (see section 2.2).

[40] The literature has largely assumed that consumers have no commitment power (see Section 2.6 for exceptions); in most contexts this assumption is both natural and also probably unimportant.

[41] [note k88] Even if all consumers have the same switching costs, if an entrant's production cost combined with consumers' switching cost exceeds the incumbent's production cost, the entrant will charge a smaller markup than the incumbent to the incumbent's old consumers and sell to some of them in a quantity-competition model, thus inducing inefficient switching (Klemperer (1988)). This result is exactly analogous to the standard oligopoly result that a higher-cost firm charges a smaller markup and so wins a socially excessive market share. It does not depend on whether or not the model allows price discrimination.

2.4.2 Multiproduct Firms

A consumer who buys several related products in a single period may incur additional "shopping costs" for each additional supplier used. These shopping costs may be the same as the switching costs incurred by consumers who change suppliers between periods. However, the dynamic issues that switching-cost models usually emphasise no longer arise. In particular, firms and consumers can contract on all prices, so the analogy with economies of scope in production is particularly strong. Thus shopping costs provide an efficiency reason for multiproduct firms just as economies of scope in production do.

The analogy is not perfect, because switching costs and shopping costs are based on specific consumer-firm matches, whereas the production-side economies of scope emphasised by Panzar and Willig (1981) and others depend only on the total numbers of sales a firm makes of each product and not on whether the same consumers are buying the firm's different products or whether some consumers use multiple suppliers.[42]

However, the analogy is particularly good if firms' product lines are sufficiently broad that most consumers use just one supplier. For example, Klemperer and Padilla (1997) demonstrate that selling an additional product can provide strategic benefits for a firm in the markets for its current products if consumers have shopping costs of using additional suppliers (because selling an extra variety can attract demand from rival suppliers for this firm's *other* varieties). This parallels Bulow et al (1985)'s demonstration of the same result if consumers' shopping costs are replaced by production-side economies of scope (because selling an additional variety lowers the firm's marginal costs of its current products). In both cases each firm, and therefore the market, may therefore provide too many different products. More obviously, mergers can be explained either by consumer switching costs (Klemperer and Padilla (1997)) or by production economies of scope.

Some results about single-product competition over many periods with switching costs carry over to multiproduct competition in a single period with shopping costs. For example, we suggested earlier in this section that oligopolists might benefit by synchronizing their sales to minimize switching and so reduce the pool of highly price-sensitive (no-switching cost) customers. Likewise multiproduct firms competing in a single period may have a joint incentive to minimize the number of consumers who buy from more than one firm, and Klemperer (1992) shows that firms may inefficiently offer similar products to each other for this reason. Taken together with the previous paragraph's result, this suggests that each firm may produce too many products, but that there may nevertheless be too little variety produced by the industry as a whole.

Finally, an important set of shopping-cost models are the "mix-and-match" models pioneered by Matutes and Regibeau (1988), Economides (1989) and Einhorn (1992). Most of this literature takes each firm's product-line as given, and asks whether firms prefer to be compatible (no shopping costs) or incompatible (infinite shopping costs); see Section 2.6.

Since shopping costs are not technologically distinct from switching costs (the term "shopping costs" merely seems more natural in a single-period context), we henceforth use the term switching costs to cover all these costs.

2.5 Entry

We have seen that switching costs may affect pricing patterns without disturbing efficiency (section 2.2), and that although they seem likely to raise prices in competition among a fixed set of oligopolists, this is not necessarily the case. This suggests that the most important effects of switching costs on competition overall may be effects on entry.

[42] As we noted in Section 2.1, if firms can discriminate between consumers, then each consumer becomes an independent market which, in the presence of switching costs, is closely analogous to a market with production economies of scope.

We will see that with real, exogenous switching costs, small scale entry to win new, unattached, consumers is often easy and indeed often too easy, but winning even part of the business of old "locked-in" customers may not just be hard, but also be too hard from the social standpoint. Furthermore, these results take the switching costs as given. Firms may also create unnecessary switching costs purely in order to discourage entry.

2.5.1 Small-Scale Entry is (Too) Easy

We saw in Section 2.3 that if firms cannot discriminate between old and new consumers, then the "fat cat" effect may make small scale entry very easy: incumbent firms' desire to extract profits from their old customers creates a price umbrella under which entrants can profitably win new unattached customers. And even after entry has occurred, the erstwhile incumbent(s) will continue to charge higher prices than the entrant, and lose market share to the entrant so long as they remain "fatter" firms with more old consumers to exploit.

So even an entrant that is somewhat less efficient than the incumbent(s) can enter successfully at a small scale that attracts only unattached buyers.

2.5.2 Large Scale Entry is (Too) Hard

But while new entrants have an advantage in competing for new customers (at least in simple models in which the "fat cat" effects operates), it is very hard for them to compete for old customers who are already attached to an incumbent, and any consumers who switch are likely to be less loyal, hence less valuable, ones. So entry may be hard if small-scale entry is impractical, due perhaps to economies of scale, or to network effects. Furthermore, even new consumers may be reluctant to invest in a relationship with a new supplier if they know that the supplier can only survive at a large scale, since in a market with switching costs consumers care about the future prospects of the firms they deal with.

Of course, this does not imply that there is necessarily too little large-scale entry. If switching costs are social costs, then large-scale entry may not be efficient even if the entrant's costs are modestly lower than an incumbent's.[43]

One argument why entry may be too hard, that is, why efficient entry may be prevented, arises if the entrant cannot discriminate between consumers. Then large-scale entry requires charging *all* consumers a price equal to the incumbent's price less the marginal old buyer's switching cost. But socially the switching cost applies only to the *old* switching buyers, not to the new consumers, and only applies to switching buyers at the average level of their switching cost, not at the marginal switching cost. So the social returns to entry are not fully captured by the entrant, and efficient entry may be blocked.

Furthermore, efficient entry can sometimes be strategically blockaded. In particular, an incumbent may "limit price", that is, cut price to increase output prior to threatened entry, to lock in more customers and make entry unprofitable at the necessary scale, when entry at the same scale would have been profitable, and perhaps efficient, if the additional customers had not been "locked-up" prior to entry (see Klemperer (1987c)).

Of course, new entry can be too easy or too hard for more standard reasons. New entry can be too hard if it expands market output, and consumers rather than the entrant capture the surplus generated. And entry is too easy if its main effect is to shift profits from the incumbent to the entrant.[44]

[43] Similarly, low cost incompatible entry can be inefficient in the presence of network effects.

[44] Klemperer (1998) illustrates the latter case, showing that new entry into a mature market with switching costs can sometimes be socially undesirable. The point is that just as entry of a firm whose costs exceed the incumbent's is often inefficient in a standard Cournot model without switching costs (Bulow et al, 1985, section, VI E, Mankiw and Whinston, 1986) so entry of a firm whose production cost *plus* consumers' switching cost exceeds the incumbent's production cost is often

But despite these caveats that apply whether or not there are switching costs, the arguments specific to switching costs suggest that entry that depends for its success on consumers switching is not just hard, but too hard.

2.5.3 Single-Product Entry May Be (Too) Hard

We saw in Section 2.4.2 that switching costs (or shopping costs) can "tie" sales together so consumers prefer not to patronise more than one firm, and it follows that a new entrant may be forced to offer a full range of products to attract new customers (let alone any old consumers). If offering a full range is impractical, entry can effectively be foreclosed. Thus in Whinston (1990) and Klemperer and Padilla (1997), the tying of a product to a product in a different market can foreclose firms that can only sell single products. In Whinston the "switching costs" are contractual, while in Klemperer and Padilla the products are "tied" by real switching costs.[45] If the switching costs are real, entry need not be too hard *given* the switching costs, but the arguments of the previous subsection suggest it often may be.

2.5.4 Artificial Switching Costs Make Entry (Too) Hard

The previous discussion addressed whether entry is too easy or too hard, taking the switching costs as given. However, the larger issue is probably whether the switching costs either are, or need to be, real social costs. They may instead be contractual such as those imposed by "loyalty contracts" that return a fraction of past payments to customers who continue to patronise the firm, or by "exclusive contracts" that require customers to pay damages if they fail to do so, or by "bundling" or "tying" products to make it uneconomical for consumers to buy single products from different firms. Or the switching costs may be real but caused by an unnecessary technological choice that an entrant cannot copy. In these cases, large-scale entry is probably too hard (see Section 2.5.2), [46] but it is the incumbent's ability to choose incompatibility that is the crucial entry barrier.[47]

2.6 Endogenising Switching Costs

To the extent that switching costs create inefficiencies (including the social cost of the switching cost itself) market participants have incentives to avoid and reduce switching costs. On the other hand oligopolists have incentives to create switching costs where they enhance market power, and a monopolist may wish to create switching costs either to deter new entry, or to extract returns from a new entrant. We take these cases in turn.

2.6.1 Reducing Switching Costs to Enhance Efficiency

As we have seen, a firm that cannot commit not to exploit its ex-post monopoly power must charge a lower introductory price. If the price-path (or quality-path) is very inefficient for the firm and consumers jointly, the firm's surplus as well as joint surplus may be increased by nullifying the switching costs. Thus, for example, a company may license a second source to create a future competitor to which consumers can costlessly switch (Farrell and Gallini (1988)).[48]

Likewise, firms producing differentiated products (or product lines) may deliberately make them compatible (i.e., choose zero switching costs). This increases the variety of options available to consumers

inefficient in a quantity-setting model with switching costs (see note K88).

[45] Choi (1996) shows that the tying of markets where R&D is critical can allow a firm with an R&D lead in just one market to pre-empt both. The welfare effects are ambiguous.

[46] Although not necessarily too hard taking the switching costs as given.

[47] [Does Farrell-Shapiro (1988) add more to this section than I've said in Section 2.3?]

[48] In Gilbert and Klemperer (2000) a firm precommits to low prices that will result in rationing but will not fully exploit the consumers ex-post, to induce them to pay the start-up costs of switching to the firm.

who can then "mix-and-match" products from more than one firm without paying a switching cost. So eliminating switching costs can raise all firms' demands, and hence all firms' profits.[49]

Where suppliers are unwilling to reduce switching costs (see below), third parties may supply converters,[50] or regulators may intervene.

We have also already noted that customers may incur the switching (or start-up) cost of using more than one supplier,[51] or may pre-commit to ignoring the switching costs in deciding whether to switch,[52] in order to force suppliers to behave more competitively.

Finally, firms may be able to mitigate the inefficiencies of distorted prices and/or qualities by developing reputations for behaving as if there were no switching costs.[53]

2.6.2 Increasing Switching Costs to Enhance Efficiency

Firms may also mitigate the inefficiencies of distorted prices and qualities by vertically integrating with their customers.[54] Likewise Taylor (1998) finds firms might set lower prices to loyal consumers to reduce inefficient switching. Of course, a downside of these strategies of increasing switching costs is that they also limit the variety available to consumers unless they pay the switching costs.

2.6.3 Increasing Switching Costs to Enhance Oligopoly Power

Although switching costs reduce social surplus, we saw in Sections 2.2 and 2.3 that they can nevertheless increase firms' profits if consumers' desire for variety (i.e., to use more than one supplier) is not too significant.[55] If so, firms may artificially create or increase switching costs. Of course, a firm may prefer switching costs *from* but not *to* its product if it can achieve this, especially where the switching costs are real social costs, but Koh (1993) analyses a model in which each duopolist chooses a real social cost of switching to it, and shows the possibility that each chooses a positive switching cost in order to relax competition.

In Banerjee and Summers (1987) and Caminal and Matutes (1990) firms have the option to generate contractual switching costs by committing in period zero to offering repeat-purchase coupons in a two-period duopoly, and both firms (independently) take this option.

And our discussion in Sections 2.2 and 2.3 suggests a slight presumption that duopolists would jointly prefer to commit to real social switching costs than to no switching costs before they compete, although the opposite is also possible.

2.6.4 Reducing Switching Costs to Enhance Oligopoly Power

An important class of models which suggests that firms may often be biased towards too much compatibility from the social viewpoint is the "mix-and-match" models (see Section 2.4) in which different firms have

[49] See Matutes and Regibeau (1988), Economides (1989), Marinoso (1998), Stahl (1982), etc. But the mix-and-match models reveal other effects too; see Section 2.6.4.

[50] [reference? discussion? see section 3.X]

[51] Greenstein (1992) discusses the procurement strategies employed by U.S. federal agencies in the late 1970s to force suppliers of mainframe computers to make their systems compatible with those of their rivals.

[52] See Cabral and Greenstein (1990).

[53] Perhaps more plausibly firms may develop reputations for, or otherwise commit to, treating old and new customers alike (since this behaviour is easy for consumers to understand and monitor); this behaviour may also mitigate the inefficiencies due to the distorted prices.

[54] See Williamson (1975) and Klein, Crawford, and Alchian (1978).

[Note FS] However *incomplete* contracts to protect against suppliers' opportunism may be less desirable than no mitigation at all of the effects of switching costs. Farrell and Shapiro (1989) call this the *Principle of Negative Protection*. The point is that it is better for customers to be exploited efficiently than inefficiently ex-post. So if contracts cannot set all future variables (e.g. can set prices but not qualities), so customers anyway expect to be exploited ex-post, it may be better that there are no contracts.

[55] [?make this caveat clearer earlier] TO DO

different abilities in producing the different components of a "system". Consumers' ability to mix-and-match the best product(s) offered by each firm is an efficiency gain from compatibility, but firms' private gains from compatibility may be even greater:

Einhorn (1992) found in a simple Bertrand duopoly model that firms (jointly) *more than* appropriate the efficiency gain from compatibility (that is, from zero rather than infinite shopping costs) and so are biased towards *excessive* compatibility. The reason is that when different firms are best at providing different components, and a single consumer wants one each of a list of components produced by both firms, the winning seller on each component appropriates its full efficiency margin in compatible competition. But in *in*compatible competition the winner's margin is its efficiency advantage where it is best, *minus* its rival's advantage where its rival is best. Conversely, the consumer prefers incompatible competition because with compatibility he receives total surplus corresponding to buying the worst of each component at cost, while with incompatibility he receives a total surplus corresponding to buying the worst bundle at cost.

Of course, this depends on (among other assumptions) duopoly at each level. With $m > 2$ firms producing each component, combining the second-best of m firms' offerings of good X with the second-best of m firms' offerings of good Y can easily give consumers *more* surplus than the second-best of m bundled offers. Farrell, Monroe and Saloner (1998) show that with $m > 2$ firms, consumers often prefer compatibility and that firms' incentives may be biased either way. Thus, in this area, assuming duopoly for simplicity is perilous.[56]

The "order-statistic" effect emphasised in these models is not the only force, however. Matutes and Regibeau (1988) stressed that under compatibility a price cut by one firm in one component increases the demand for the *other* firms' complements, so that compatibility reduces incentives to cut prices.[57] Economides (1989) argued that, unlike the Einhorn result, this logic does not depend on duopoly, so provides a clear argument why firms may try too hard to reduce switching costs and shopping costs.[58]

2.6.5 Increasing Switching Costs to Prevent or Exploit Entry

We have seen that an incumbent firm may protect a monopoly position against entry by writing exclusionary contracts, or by artificially creating real switching costs through technological incompatibility with potential entrants.[59] Imposing contractual switching costs (but not real social switching costs) can also be a mechanism for extracting rents from an entrant without preventing its entry—the entrant is forced to pay a fee (the "liquidated damages") to break the contracts.[60]

[56]Likewise, Einhorn's results, but not those of Farrell, Monroe and Saloner, are qualitatively unaffected by whether or not firms know their own efficiencies in each component.

[57]Matutes and Regibeau (1992) allowed firms to set separate prices for bundles (not necessarily the sum of the component prices) and found that the force toward compatibility weakens. Furthermore, compatibility also changes the structure of demand, so even Matutes and Regibeau (1988) found that firms are sometimes on net biased towards incompatibility. And Klemperer (1992) also shows that firms may prefer incompatibility to compatibility when the latter is socially preferred, and that the firms may even distort their product choices to sustain incompatibility. (All these models assume some product differentiation between firms' components even under compatibility). See also Anderson and Leruth (1991) and DeNicolo (2000).

[58]Most of the "mix-and-match" literature assumes that each firm offers a full line of products, but DeNicolo (2000) analyzes competition with one full-line and a pair of specialist firms. So, in our terminology, there are then no additional shopping costs of buying from an additional specialist firm after having bought from one of the specialist firms, but the specialist firms do not internalize the complementarities between them.

[59]Imposing switching costs would not be worthwhile for the incumbent if they reduced consumers' willingnesses to pay by more than the gains from excluding entry. In Rasmusen, Ramseyer, and Wiley (1989) entry is only possible above some minimum efficient scale that is greater than any single customer's demands, so no individual customer loses by forgoing the possibility of using an alternative supplier if other customers have already forgone the possibility, and so no customer needs to be compensated for signing an exclusive contract.

Deterring entry is also profitable if it can transfer rents from an entrant to the incumbent.

[60]See Aghion and Bolton (1987).

2.7 Battles for Market Share

2.7.1 The Value of Market Share

With switching costs (or indeed proprietary network effects—see Section 3.7), a firm's current customer base is an important determinant of its future profits.

We can therefore write a firm's current-period value function, (i.e., total discounted future profits), V_t, as the sum of its current profits, π_t, and its discounted next-period value function $\delta V_{t+1}(\sigma_t)$ in which δ is the discount factor, and the next period function value, $V_{t+1}(\cdot)$, is a function of the size of its current-period customer base, σ_t.

$$V_t = \pi_t + \delta V_{t+1}(\sigma_t) \tag{1}$$

Obviously this is a strong simplification; firm's future profits may depend on which customers it has sold to, and these customers' full histories, how market share is distributed among competing firms, etc. However, in the simplest models it is equivalent to let V_{t+1} be a function of current-period market share—for example, this is true of Klemperer (1987b, 1995) Farrell and Shapiro (1988), Beggs and Klemperer (1992), Padilla (1992, 1995), ?? Chen and Rosenthal (1985)[61] which are all models with just two firms and a fixed set of consumers whose reservation prices are sufficient that they always purchase. So σ_t is often interpreted as "market share", and this explains the very strong concern with market shares in markets with switching costs and/or network effects.[62]

2.7.2 Penetration Pricing

When the firm sets a single price (as, for example, in our discussion of "the level of prices" in Section 2.3) its first-order condition is

$$0 = \frac{\partial V_t}{\partial p_t} = \frac{\partial \pi_t}{\partial p_t} + \delta \frac{\partial V_{t+1}}{\partial \sigma_t} \frac{\partial \sigma_t}{\partial p_t} \tag{2}$$

Provided that a firm's value fraction is increasing in its customer base, $\partial V_{t+1}/\partial \sigma_t > 0$,[63] therefore, the firm charges a lower price or sets a higher quantity than would be short-run profit maximising in order to raise its customer base and hence its future profits. That is, $\partial \pi_t/\partial p_t > 0$ (since we assume $\partial \sigma_t/\partial p_t < 0$).

In the early stages of a market, therefore, when even short-run profit-maximising prices are not high relative to costs, equation (2) implies low, "penetration", pricing. This is the familiar story behind, for example, banks offering gifts to induce customers to open new accounts (because of switching costs) the extreme version of which we have already described in Section 2.2.

The larger the value of the future market, V_{t+1}, the deeper, ceteris paribus, the penetration pricing will be. For example, a more rapidly growing market will have lower prices.[64] [65]

[61] [CHECK] [OTHERS?]

[62] Because switching costs make current market share such an important determinant of a manufacturer's future profits, Valletti (?) suggests they may provide a strong motive for vertical integration with retailers to ensure sufficient investment in a base of repeat subscribers.

[63] This need not apply: stealing customers from rival(s) may make the rival(s) so much more aggressive that the firm is worse off. See Banerjee and Summers (1987), Klemperer (1987c). Usually, however, $\partial V_{t+1}/\partial \sigma_t > 0$. In Beggs and Klemperer (1992), V_{t+1} is convex (and quadratic) in σ_t. (The fact that the sum of the duopolists' value functions is therefore maximised at the boundaries is consistent with stable dynamics because lowering current price is less costly in current profits for the firm with the smaller market share.)

We can perform a similar analysis with similar results for a quantity-setting firm.

[64] See, for example, Beggs and Klemperer (1992), Borenstein, Mackie-Mason and Netz (2000) and also Holmes' (1990) steady-state model of a monopolist selling a single product to overlapping generations of consumers who incur set-up costs before purchasing the product.

[65] It seems unclear whether we should expect "penetration pricing" patterns from a monopolist, since the magnitude of $\partial V_{t+1}/\partial \sigma_t$ may be smaller in monopoly—where consumers have nowhere else to go—than in oligopoly, and durable-goods effects imply falling prices in monopoly absent switching-cost effects (equation (2) only implies that early period prices are lower than in the absence of switching-costs, not that prices necessarily rise). Cabral et al show it is hard to obtain penetration pricing in a network-effects monopoly model (see Section 3.6).

Of course, in a more general model, the "penetration" may be through advertising or other marketing activities rather than just low prices. This is especially likely when it is hard to price far below cost (for example, it may be hard to charge negative prices without attracting worthless customers who are not real users or repeat buyers).

2.7.3 Harvesting vs Investing: Macroeconomic and International Trade Applications

Any firm must in every period balance the incentive to charge a high price to "harvest" greater current profits against the incentive for a low price that "invests" in market share (or customer base)and hence increases future profits—see our discussion of equation (2).

Anything that increases the marginal value of market share will make the firm lower price further to invest more in market share. Thus, for example, equations (1) and (2) shows that a lower δ, that is, a higher real interest rate reduces the present value of future market share so leads to higher current prices.[66]

Chevalier and Sharfstein (1996) develop this logic in a switching-costs model based on Klemperer (1995). They argue that liquidity-constrained supermarkets perceive very high real interest rates and therefore set high prices, sacrificing future profits in order to raise cash and survive in the short term. Their theory is supported by empirical evidence that shows that the most financially-constrained supermarket chains do indeed raise their prices relative to less financially-constrained chains during recessions.[67],[68]

Fitoussi and Phelps (1988) argue that the same effect helps explain the high rates of inflation in Europe in the early 1980's, although these authors emphasise search costs rather than switching costs.

Froot and Klemperer (1989) apply the same logic to international trade in a general model of competition for market share motivated by switching costs and network effects. An appreciation of the domestic currency lowers a foreign firm's costs (expressed in the domestic currency) so tends to lower prices, but if the appreciation is expected to be only temporary then the fact that the domestic currency will be worth less tomorrow is equivalent to an increase in the real interest rates which raises prices. So exchange-rate changes that are expected to be temporary may have very little impact on import prices. Conversely, if the currency is anticipated to appreciate in the future, both the "interest-rate effect" and "cost effect" are in the same direction—market share tomorrow is probably worth more if future costs are lower, and tomorrow's profits are worth more than today's profits, so for both reasons today is a good time to invest in market share rather than harvest current profits. So import prices may be very sensitive to anticipated exchange-rate changes. Both Froot and Klemperer (1989) and Sapir and Sekkat (1993) provide empirical support for these theories.[69]

2.8 Competition Policy

As we have seen, with (large) switching costs firms compete over streams of goods and services rather than over single transactions. So one must not jump from the fact that buyers become locked in to the conclusion that there is an overall competitive problem. Nor should one draw naive inferences from individual transaction prices, as if each transaction were the locus of ordinary competition. Some individual transactions may be priced well above cost even when no firm has market power; others may be priced

[66] All these arguments are discussed in more detail in Klemperer (1995).

[67] [Cite Borenstein?]

[68] Beggs and Klemperer (1989, Section 5.3) and Klemperer (1995) provide further discussion of how "booms" and "busts" affect the trade-offs embodied in equation (2) and hence affect price-cost margins.

[69] For other applications of switching-costs theory to international trade, see Tivig (Industrial Journal of Industrial Organisation 1996, p631) who develops "J-curves" (since sales quantities respond only slowly to price changes if there are switching costs) To (1994), and Hartigan (1995).

below cost without being in the least predatory.[70] [71]Thus switching-cost markets can be more competitive than they look, and switching costs need not generate supernormal profits, even in a closed oligopoly.

But while switching costs need not cause competitive problems, they probably do make competition more fragile, especially when they coexist with ordinary scale economies. Because large-scale entry into switching-cost markets is hard, there may be much more incentive for monopolizing strategies such as predation or merger than there is in markets in which easy entry limits any market power. Thus switching costs, in combination with other factors, could justify heightened antitrust scrutiny.

Furthermore, while they need not, switching costs often do raise average prices, and often cause an inefficient bargain-then-ripoff pattern of prices even when the average level of prices remains competitive; they make matching less efficient by discouraging re-matching or the use of multiple suppliers; and, of course, they result in direct costs when consumers do switch. Perhaps most important, while sometimes (as in simple bargain-then-ripoff models) firms must give all their ex post rents to consumers in ex ante competition, that is not always true. The ex post rents may be less than fully competed away[72], or they may be dissipated in unproductive activities such as excessive marketing or advertising in which case consumers are harmed by switching costs, even though firms may be no better off.[73]

For these reasons, despite the warnings in our first paragraph, markets may indeed perform less well with switching costs than without, so policy intervention to reduce switching costs may be appropriate.[74] For example, policy might cautiously override intellectual property rights, especially of copyright-like intellectual property that may have little inherent novelty, if those rights are used only as a tool to enforce incompatibility and so create private rewards that bear no relationship to the innovation's incremental value.

In general firms may be biased either towards or against compatibility relative to the social standpoint. But switching costs seem more likely to lower than to raise efficiency, so when firms favor switching costs the reason is often because they enhance monopoly or oligopoly power by directly raising prices or by inhibiting new entry.[75] This suggests that policy-makers should take a close look when firms with market power choose to have switching costs (through contract form or product design) when choosing compatibility would be no more costly.[76] [77]

---

[70]For instance, in an aftermarket context such as the Kodak case, the fact that repair services are priced well above cost does not by itself prove that there is a serious competitive problem.

[71]Another naïve argument is that if one observes little or no switching, then firms do not constrain one another's prices: firms that compete on a life-cycle basis (rather than on an individual transaction basis) constrain one another's life-cycle prices and, of course, firms may be constrained even ex post by the threat of customer switching even when that threat is not carried out in equilibrium.

[72]There seems to be a perception that this might often be the case, see for example Porter (1980 p19, p120) [SAY MORE]. It is hard to give many good reasons why, but possible reasons are that firms may be liquidity constrained and/or risk-averse, ex-ante, or that a firm (or collusive group of firms) with a first-mover advantage faces myopic consumers.

[73]Firms often pay consumers' switching costs, but this does not eliminate the competitive significance of those costs. In fact, one can read the simple model of e.g. section 2.3.1 as predicting that "the" price will be $c + (1 - \phi\delta)s$ and firms "pay the switching (or setup) costs."

[74]Gans and King (2000) examine the regulatory trade-offs in intervening to reduce switching costs and show that who is required to bear the costs of ameliorating switching costs may be an important determinant of the efficiency of the outcome. Aoki and Small (1999) analyse number portability in the telecoms market and Galbi (2001) describes tradeoffs in setting switching costs in long-distance telecommunications.

[75]A caveat is that firms often do not make a coordinated joint choice of whether to compete with switching costs or without, and different firms may be able to control the costs of different switches. See Section 2.6.

[76]For example, the Swedish competition authority argued that Scandinavian Airlines' "frequent-flyer" program blocked new entry on just one or a few routes in the Swedish domestic air-travel market in which entry on the whole range of routes was impractical, and the airline was ordered to alter the program from October 2001. See Section 2.5.3. (For more on frequent-flyer programs see Klemperer and Png, 1986.)

[77]A caveat is that the policy debate is often held ex-post of some lock-in. At this point incumbents obviously favor maintaining high switching costs and their preference is not evidence that switching costs raise prices overall (nor is it necessarily inefficient). Reducing switching costs ex-post also expropriates the incumbents' ex-ante investments, possibly raising concerns about harming dynamic efficiency.

3 Network Effects and Competition

3.1 Introduction

When scarcity drives prices, it pays to be different and avoid the crowd. At a given income, people who live in South Dakota can afford more books than those who live in Southern California.

In other contexts, it pays to coordinate and follow the crowd. A human language is more useful when more people speak it. Driving is easier if everyone keeps right — or if everyone keeps left. A telecommunications network with more users gives each of them a wider option to communicate. A larger market may have "thicker" and thus more efficient trade. Such benefits are *network effects*. A network effect is *complementarity* between my adoption of a good and yours: network effects stem from a consumer's desire for compatibility with other consumers' choices. Additional adoption both makes existing adopters better off (a total effect) and increases the incentive to adopt (a marginal effect).

Efficiencies of market size can create network effects, just as efficiencies of repeated dealing create switching costs.[78]

The examples above display technological complementarities: a "direct" network effect. An "indirect" network effect arises when wider adoption benefits adopters by changing the behavior of other actors — sellers of that good, or sellers or buyers of a related good. In particular, indirect network effects arise when widespread adoption improves the supply of that good or of a complement: when a bigger market benefits buyers.[79] Thus, scale often increases incentives for innovation and investment in product quality. It might induce more product variety, as in matching markets. Computer hardware is more valuable the better the selection of software available to run on it, and that selection will be broader for more popular hardware. Likewise, a computer operating system gains from a wide variety of applications software.

Network effects are often externalities: one person's adoption decision affects another's payoff. If this effect is indeed an externality, a network good will be under-adopted. Early literature focused largely on this topic, and in particular on how a regulated telecommunications monopoly should price subscriptions.[80] But simply internalizing the network effect at the margin leaves some tough problems of coordination. In particular, there may be multiple equilibria among adopters, making coordination a challenge and expectations very important. As a result, network markets often display unstable dynamics (critical mass and tipping) and path dependence, including a kind of collective switching costs, so penetration pricing and bargain-then-ripoff pricing are natural strategies.

For the same coordination reasons, if networks compete with one another and if adopters will not smoothly coordinate on the best deal, a focal product may enjoy a great deal of market power. Business strategy thus involves helping buyers and complementors to coordinate on one's product, and perhaps hindering them from coordinating on rivals'. Correspondingly, competition policy takes a strong interest in network effects. In 1998 Microsoft was accused of illegal actions aimed at maintaining the "applications barrier to entry:" the self-reinforcing tendency of writers of application software, and other complementors, to make it their first priority to work well with the dominant operating system.[81] This indirect network

---

[78] Market size can also shift terms of trade, creating a *pecuniary* network effect, as we will discuss below.

[79] Explicit theories of indirect network effects created through improved supply in a complement are in Church and Gandal (1992, 1993) and Chou and Shy (1990). This literature is reviewed by Gandal (1995) and the topic is helpfully revisited by Katz and Shapiro (1994).

On incentives for mix-and-match compatibility (often thought of as network effects), see section 2.6.4 above.

[80] See e.g. Squire (1973), Rohlfs (1974), Kahn and Shew (1987).

[81] Both the Department of Justice and Microsoft have made many documents available on their web sites, www.usdoj.gov/atr/ and www.microsoft.com/ respectively. An introduction to the case is the 2001 decision of the DC Court of Appeals. An extended discussion by economists involved in the case is Evans *et al.* (2000); Fisher (2000) and Schmalensee (2000) give briefer discussions; see also Gilbert and Katz (2001), Evans and Schmalensee (NBER 2001), and Bresnahan (2001). Werden (2001) focuses on the applications barrier to entry. (Farrell was a consultant, and later Deputy Assistant Attorney General, for the Justice Department.)

effect is at the heart of the case.

While most of the literature is theoretical, as is this chapter, there has been both historical and econometric work on network effects and standards.[82]

In this section, we analyze network effects and their consequences for competition. For the most part, like the literature, we do not focus explicitly on whether the effects are direct or indirect.[83] Section 3.2 discusses the possible externality. Section 3.3 introduces the coordination problem; section 3.4 discusses institutions for coordination, and section 3.5 discusses inertia in sequential adoption. Those sections describe coordination of adoption given prices; we then turn to price-setting. Section 3.6 introduces monopoly pricing and section 3.7 concerns competitive pricing with incompatible products, and implications for innovation. Section 3.8 brings together discussion of firms' incentives to be compatible or incompatible, and section 3.9 discusses public policy on compatibility.

3.2 Network Effects and Externalities

Consider K players, whom we often call adopters or buyers. Each chooses an action: to adopt a product or not, or to adopt one product or another. Player i has payoff $u_a^i(x)$ from action a if the players who do a amount (in number or measure) to a total of x. There are network effects in a if, for each i, $u_a^i(x)$ is increasing in x.[84]

For simplicity, like much of the literature, we focus largely on the simple case where $K = 2$. We often interpret these players not as individuals but as "groups", so we specify that i is of size n_i and $n_1 + n_2 = N$. We follow most of the literature in assuming away any *within*-group externalities or coordination problems and focus on such problems *between* groups.[85]

To begin, consider the single-product case: players choose whether or not to adopt a single product. We can then normalize each player's payoff from non-adoption as zero (the alternative is not a network good, so this payoff is unaffected by others' adoption choices), and (dropping the action subscript) write $u^i(x)$ for player i's payoff from adoption if total adoption amounts to x. These payoffs are illustrated in *Figure 1*:

	2 does not adopt	2 adopts
1 does not adopt	$0, 0$	$0, u^2(n_2)$
1 adopts	$u^1(n_1), 0$	$u^1(N), u^2(N)$

Network effects are then present if $u^i(N) > u^i(n_i)$ for $i = 1, 2$. We describe the natural extension to the case of competing network products in 3.3.2 below.

[82] Historical studies include David (1985) on the QWERY typewriter keyboard and the responses by Liebowitz and Margolis (e.g. 2000); Cusumano, Mylonadis and Rosenbloom on VCRs (Bus Hist Rev 1992), Thompson (19xx) on ball bearings. Bresnahan (2001) compiles some of Microsoft's internal discussions on network effects and business strategy.

Econometric work includes Gandal (1995) on PC spreadsheets, Saloner and Shepard (1995) on ATMs, Brynjolfsson and Kemerer (1996), Gandal, Kende and Rob (2000) on compact disk players, Dranove and Gandal (2001) on DVDs, Goolsbee and Klenow (2000) on personal computers, Park (2001), and Rysman (2000) on the Yellow Pages.

[83] Economides (1996) calls this the "macro" approach, in contrast to the "micro" approach of modeling the complementarities in full.

[84] Except where one good has no network effects, we generally assume for clarity that this increasing property is strict. A more general formulation would allow each user i to gain more from the presence of one other user j than of another k. For instance, each telephone subscriber gains more network benefit from adoption by her friends than from adoption by strangers, even if the latter would use the phone just as much: Beige (2001) explores such "local" network effects. But the simpler formulation in terms of network "size" dominates the literature; Segal (1999) shows that this (his Condition S) leads to substantive as well as expository simplifications. On the other hand, a more special formulation than the text's assumes that $u_a^i(x)$ is linear in x and independent of i, so that the total value of the network is quadratic in x: a claim known as Metcalfe's law.

Here and elsewhere, we distinguish between x and n: x is the total adoption of a network, while n_i is the size of group i.

[85] In principle, one can maintain this assumption by declining to aggregate into a "group" players who do not coordinate well together. However, the literature has not always followed such a prescription.

We say that network effects are *strong* if each player would like to do whatever the other does: coordination outweighs individual differences in tastes. Here, the condition is that $u^i(N) \geq 0$ for all i. We say further that preferences are *similar* if the players agree on the best outcome. Since each player wants everyone else to do whatever he does, here, the condition is that $u^i(N)$ has the same sign for all i.[86] Observe that when preferences are similar, the efficient outcome is an equilibrium of the simple adoption game suggested by Figure 1.

3.2.1 Are Network Effects Externalities?

Network effects are often externalities, but need not be. Liebowitz and Margolis (e.g. 1994) have stressed that some literature too cavalierly assumes that network effects are real externalities, in two respects.

First, any economic effect is an externality only if it is not internalized. There could be (and there are incentives to develop) contractual or other ways to internalize a network effect, through side payments between adopters or through pricing by a seller: an unregulated seller may voluntarily subsidize marginal adopters if it can capture the increase in inframarginal adopters' willingness to pay; see 3.6.1 below.

Second, network effects can be *pecuniary*: terms of trade, as well as efficiency, may shift as a market expands.[87] Then, although Figure 1 describes payoffs to buyers, sellers bear a corresponding negative effect.[88] We will assume that network effects are real (at least not *purely* pecuniary) efficiencies of coordination.[89]

3.2.2 Under-Adoption of Single Network

When network effects are externalities, textbook economics would suggest generic under-adoption of a single network good, in the sense that the efficient outcome fails to be an equilibrium. As we noted above, the efficient outcome *is* an equilibrium if adopters have similar preferences. But, if they do not, then efficient adoption can unravel: a few are unwilling to adopt even if all others do so; their defection makes others less keen and leads a few more to defect, and so on: Schelling (1978) describes such quasi-dynamics.

An important special case arises if a network good is *competitively supplied.* Then players who value the good (given anticipated network size) at less than marginal cost will not adopt. Because of the externality, some of them should: generically the network will be too *small*;[90] in corresponding dynamic terms, given a competitive price path, adoption will tend to be too *slow*.[91] Optimality requires subsidizing the marginal adopter to the extent of his external contribution to others.[92]

---

[86] With two network goods, as we will see, strong network effects and similar preferences are less closely related than here.

[87] Liebowitz and Margolis note that if a good has perfect competition and (industry-level) decreasing costs, so that the price falls with the quantity demanded, then there is a pecuniary network effect among buyers, with an opposite effect on their immediate suppliers. But this may not complete the welfare accounting: such competitive decreasing costs often rest on a real efficiency of scale upstream.

[88] This applies with simple modification also to pecuniary indirect network effects, for instance where it is a complementary good for which the terms of trade shift. Liebowitz and Margolis (199x) indeed claim that almost all indirect network effects are pecuniary.

[89] Purely contractual direct network effects (such as MCI's "Friends and Family" program) seem less widespread than contractual switching costs. Indirect proprietary network effects can be generated by exclusivity (see Economides and White for such analogies). In telecommunications, a pecuniary network effect can be generated if on-net calls are cheaper than off-net calls, perhaps because carriers charge one another more than marginal cost for termination.

[90] Beige (2001) shows that equilibrium locally maximizes a "harmony" function that counts only half of the network effects in the sum of payoffs.

[91] Such dynamic adoption paths have been studied by e.g. Rohlfs (1974), Farrell and Shapiro (1992, 1993), Economides and Himmelberg (1995), Choi and Thum (EER 1998), and Vettas (2000).

[92] This argument has been made to support "universal service" subsidies in telecommunications. Many economists are skeptical about this application: see recently Einhorn (RIO 1993), Yannelis (IJSocEcon 2001), and Barnett and Kaserman (J Reg Econ 1998). But Cremer (EER 2000) gives such an argument for internet broadcasting.

3.3 The Coordination Problem

Most of microeconomics, like 3.2.2 above, assesses *incentives*, usually asking whether there are incentives to deviate from an efficient outcome. Such incentives do normally obstruct efficiency; but even if an efficient outcome would be an equilibrium, coordination on it may be hard. As we will see, efficient coordination depends crucially on *expectations*, which in turn depend on *institutions* for adoption decisions.

3.3.1 Coordination with One Network Good

When a single network good has strong network effects, a simultaneous-adoption game has multiple equilibria. Since $u^i(N) > 0$ but $u^i(n_i) < 0$ for all i, there are both a no-adoption equilibrium and one in which all adopt.[93] Each group will adopt if and only if it expects enough others to adopt: *expectations are king.*

The equilibrium with the most adoption Pareto-dominates non-adoption and any other equilibria.[94] Inefficiently coordinating on an inferior (non-maximal) equilibrium is a form of under-adoption rather different from the marginal failure in incentive discussed in 3.2.3 above. A desirable technology may fail to achieve "critical mass:" a metaphor that reminds us that once enough adoption happens or is confidently foreseen, further self-reinforcing adoption follows — but by the same token lack of adoption is also self-reinforcing: a network product fizzles if too little adoption leads others to decide not to adopt (the "death spiral").[95]

3.3.2 Competing Network Goods

Often the choice is not simply between adoption and not, but between adopting one network good and another; this leads to *Figure 2*:[96]

	2 adopts A	2 adopts B
1 adopts A	$u_A^1(N), u_A^2(N)$	$u_A^1(n_1), u_B^2(n_2)$
1 adopts B	$u_B^1(n_1), u_A^2(n_2)$	$u_B^1(N), u_B^2(N)$

Network effects are then present if $u_a^i(N) > u_a^i(n_i)$ for $i = 1, 2$ and $a = A, B$. We say that network effects are "strong" if they outweigh preferences for A versus B,[97] so that "all adopt A" and "all adopt B" are both Nash equilibria of the simultaneous-move noncooperative game with payoffs as in Figure 2; for clarity we focus largely on this case. We say that players have "similar preferences" if they agree on the ranking of those two extreme outcomes.

As above, *if preferences are similar then the efficient outcome is an equilibrium.*[98] It is clearly an equilibrium of that simultaneous-move game; but a stronger statement also holds. Institutions (or things that change the game form, given primitive preferences) often change the outcome in games in general. But when, as here, an equilibrium is also each player's best feasible outcome, many institutions (including

[93] With more than two groups and non-similar preferences, it can also be a strict equilibrium for some, but not all, to adopt. There may be more than two pure-strategy adoption equilibria; there may also be mixed-strategy equilibria, which we do not pursue here.

[94] Dybvig and Spatt (1983). First, there is such a *maximal equilibrium*: adoption by all players who adopt in any equilibrium is an equilibrium. Second, adoption by one player benefits other adopters and does not harm non-adopters, so voluntary adoption is Pareto-improving. Thus the maximal equilibrium is Pareto preferred to all other equilibria. This argument rests on the positive externality and does not extend to the choice between network goods. Pareto-dominance is here assessed among the adopters. Of course, if price is at or above incremental cost, the Pareto-dominance extends to the seller.

By construction, the maximal equilibrium has the most adoption in a set-inclusion sense, as is characteristic of games with supermodularity: see Topkis (1998) or Milgrom and Roberts (1990).

[95] Leibenstein (1950) noted the multiple equilibria, including the possibility of both zero and positive adoption equilibria. Rohlfs (2001), among many others, discusses the process of getting to critical mass. Jeitschko and Taylor (AER March 2001) investigate the stability of what they call "faith-based coordination."

[96] By assuming that one of the products has zero network effects, one can subsume the former case in this one. We present both cases for expositional reasons.

[97] Formally, if $u_a^i(N) \geq u_{a'}^i(n_i)$ for $i = 1, 2$ and $a, a' = A, B$. [do we need this??]

[98] Has this observation been published?

side payments, order of moves and commitment, communication) preserve the equilibrium. There is no evident reason to use or develop an institution that will move anyone away from the best outcome.

Thus, inefficiency in adoption will normally arise only when (a) preferences differ (or may differ), and/or (b) it's hard to coordinate on the efficient equilibrium. Both (a) and (b) may well hold most of the time, but the observation helps us organize our thinking.

In contrast to the single-network case, there now are countervailing effects on others' payoffs. Adopting A, and conferring a network benefit on other A-adopters, means not adopting B, and thus withholding a benefit from B-adopters. The *net* external effect is ambiguous, and is smaller in magnitude than either $x_i u_i'(x_i)$.[99] In marginal terms, a small shift of users from a network of size x_A to one of size x_B has a net externality proportional to $e = x_B u_B'(x_B) - x_A u_A'(x_A)$: this has ambiguous sign.[100] Thus the "total effect" – that adoption makes other adopters better off – does not carry through cleanly to competing network goods.

While the externality is weaker, the marginal effect or bandwagon incentives are *stronger* than in the one-network case. When another player adopts A instead of B, it makes A more attractive and B less so. The gain in A's relative attractiveness is proportional not to $u_A'(x_A)$ as with a single network good, but to $u_A'(x_A) + u_B'(x_B)$. Thus the "marginal effect" – that adoption makes others keener to adopt – is actually strengthened with competing network goods.

Because expectations are king, it is critical to explore how expectations may be formed. As we will see, the central questions are whether expectations will reliably focus on an efficient equilibrium, and what will happen if preferences are not "similar," so that adopters do not agree on a "best" equilibrium.

3.4 Institutions For Coordination

For efficiency, adopters must coordinate well in two senses. First, they must focus on a single equilibrium, or they simply fail to coordinate. Even mere *fear of* such coordination failure can persuade adopters to *wait*, and adoption may be inefficiently slowed because of strategic uncertainty rather than because of the externality.[101] But the literature has not stressed this possibility, and we do not do so here.[102]

Second, not only must expectations focus on a single equilibrium: it must be the best equilibrium (at least when equilibria are Pareto ranked). At the level of adopters alone (treating sellers as passive for the time being), the literature has focused on two institutions that may help coordination: communication and sequential choice.

[99] In a two-group model, Farrell and Saloner (1986b) examined incentives for either group to defect from a technology chosen by the other. Either group can have an excessive group incentive to splinter but cannot have too weak an incentive to do so, since only one other group is affected; this might suggest using a three-group model (see 3.5.1 below). Farrell and Saloner instead relaxed the assumption that each group efficiently coordinates internally; then there can be excessive standardization, although not in unique equilibrium. Of course, one could phrase this as a finding about adoption with many groups each of which does coordinate internally.

[100] The incentive to defect from the mainstream is too strong at the margin (defection imposes a negative net externality) if $e < 0$ whenever $x_B < x_A$. When the goods are homogeneous except for network size, that amounts to the condition that $xu'(x)$ is increasing, or that $u(\cdot)$ has relative risk aversion $-xu'/u'' < 1$. See Farrell and Saloner (1992). In the often used and analytically convenient case $u(n) = vn$, u has zero second derivative, so there is too much incentive at the margin to defect from a network to which most players will continue to adhere. Liebowitz and Margolis (1994) stress the possibility that network effects are essentially exhausted at relevant scales, so that the u function flattens out. This is consistent with a bias, but any bias may be swamped by differences in tastes.

Goerke and Holler (Public Choice 1995), Rysman (2000), and Woeckener (1999) stress such inefficiencies of splintering.

[101] A number of models have focused on the *timing* of adoption, usually with an exogenous driver such as steadily declining costs. See for instance Farrell and Shapiro (1992, 1993), Economides and Himmelberg (1995), Vettas (2000). Generally, however, they have focused on dynamic manifestations of the externality, not on the strategic uncertainty.

Choi (1994) notes, on the other hand, that strategic adopters eager to commit may adopt too quickly when there is e.g. technological uncertainty.

[102] Mixed-strategy equilibria can model a kind of coordination failure (Dixit and Shapiro 1985, Farrell 1987, etc.).

3.4.1 Cheap Talk

A natural response to a coordination problem is to talk. Cheap talk does not solve raw incentive problems, but even cheap talk can often help players coordinate on a Pareto-preferred equilibrium. It works less well when there are many participants or where there is conflict, as there often may be with competing network products; and especially when there are both large numbers and conflict.[103] Beyond informal communication, many more or less formal standards development organizations (SDOs) bring together participants to collectively focus expectations.[104]

3.4.2 Sequential Choice Dynamics

Theory claims that fully sequential adoption prevents coordination failures. The argument is fairly convincing with two groups. For simplicity, first consider the single-network case. Suppose that $u^i(N) > 0 > u^i(n_i)$ for all i so that adoption is an efficient equilibrium and non-adoption is an inefficient equilibrium. If group 1 first adopts, then group 2 will rationally adopt: group 1 can get $u^1(N)$ by adopting and seizing control of group 2's expectations. The theoretical argument extends to any finite number of players.[105]

Unfortunately, the argument is delicate and not compelling with many players or groups: it depends on a chain of backward induction with (on the order of) K steps. If K is in the dozens, or millions, it would be a hard sell to use game-theory logic to persuade the first player that he should confidently (no matter how negative is $u(x)$ for small x) expect to be followed if he adopts. This is so even if choices will be sequential in the strict game-theoretic sense, which is not likely.[106] Again, large numbers (especially combined with conflict) tend to make optimal coordination difficult.

The result also holds for competing networks if preferences are similar. Unfortunately, again, the logic is delicate (a long chain of backward induction) when there are many players or groups; and if enough early adopters fail to see the logic or don't trust it, then it becomes wrong.

3.4.3 Rougher and readier cues for coordination

Adopters may use imperfect cues for coordination.[107] Just as some species of animals return to fixed places for breeding (presumably because coordination is more important than optimal location), adopters might, rationally or not, look to more or less arbitrary cues such as authority (what does a high-profile pundit say?), tradition, or signals such as advertising/marketing.[108] Or expectations might track quality but ignore price (in section 3.7.1 below we explain why this may be plausible). If all adopters use the same such cue, they will coordinate, but not necessarily on the best offer. When there are many players, coordination will often tend to be rather clumsy. As we will discuss below, this may lead to competitive as

[103] On cheap talk see Crawford and Sobel (1982), Farrell and Rabin (1996). Conflict is likely in technology adoption for many reasons, but in particular because of installed base (see below).

[104] Examples include the International Telecommunications Union (ITU), and a wide variety of national standards bodies such as ANSI in the US; ANSI is an umbrella organization for specialized industry standards development. There are also many informal standards fora. See e.g. Hemenway (1975), Cargill (1990), Kahin (1995), and much discussion in the business press. To a great extent their problems stem from conflict over which standard to adopt. Farrell and Saloner (1988), Farrell (1993), and David and Monroe (1997) discuss how conflict can mar the standards coordination process, using a war of attrition model; see also Bulow and Klemperer (1999). On consensus standards development see also Sirbu et al. (1991?).

[105] Farrell and Saloner (1985). The proof is an induction on K. Consider the first player who would adopt in the maximal equilibrium. If he adopts, he puts the rest of those players into a subgame whose maximal equilibrium implements the maximal equilibrium of the original game.

[106] While many real-world adoption processes are somewhat sequential (they can hardly be absolutely simultaneous), the argument depends on each player *observing all previous choices* before making his own — unlikely when there are many adopters. Nevertheless, large (groups of) adopters will recognize that they may be able to start a bandwagon in their favor: see the New Hampshire Theorem below.

[107] Imperfect coordination is discussed in Friedman (1993), Farrell (2000) and Bolton and Farrell (1990) [no doubt plenty of other places].

[108] Bagwell and Ramey (1994) explore how advertising could coordinate buyers' behavior to achieve benefits of (supply-side) economies of scale.

well as direct inefficiencies. In particular, if expectations track past success, the market will behave much like markets with switching costs even if there are no physical installed bases.

3.5 Inertia in Adoption

In many network markets, as in switching-cost markets, early adopters make complementary specific investments. Often, another (incompatible) network good later looks more attractive on a clean-slate comparison. New buyers would like everyone to adopt the new good; but the "installed base" faces switching costs and would thus prefer that others adopt the established technology. So adopters' preferences between old and new network goods are generically not "similar", even if their fundamental tastes are identical.

Will the new technology displace the old? Should it – given that switching involves real costs?[109] If the market *inefficiently* stays with the established technology and spurns the new one, we say there is *ex post excess inertia*; if it is not reluctant enough, it is *ex post excess momentum.*[110]

3.5.1 Ex Post Inertia Among Adopters

When technology B first becomes available, call the already committed installed base on A group 1.[111] Adopters immediately following B's arrival may prefer that everyone remain on A (because if all new adoption goes to B, they will suffer painful interim losses of network benefits, even though B's network may eventually grow very large); if so, call those adopters group 2. Group 3 consists of later adopters still, who prefer that everyone *starting with B's introduction* adopt B (they do not suffer from B's temporarily very small network). Within each group, preferences are similar, so suppose that in-group coordination is smooth.

Then, if group 2 is not empty, it gets its way: if it picks A, then group 3 will follow. The reason is that (at least early adopters among) group 3 will then find A still more attractive relative to B than did group 2. Since group 2 gets its way, the outcome is tilted toward group 2's preferences (sticking with A) from the joint perspective of groups 2 and 3; but this may be a bias in either direction as judged jointly by all three groups, since group 1 prefers that all stick to A.

On the other hand, if *all* adopters other than the installed base prefer that all adopt B, once it is offered, then group 2 is empty and group 3 gets its way, stranding group 1, and this may be inefficient excess momentum. There cannot be *ex post* excess inertia with only two groups: there is a *positive* externality from remaining compatible.[112]

3.5.2 Ex Ante Inertia among Adopters

Although it is simplest to describe inertia in ex post terms as above, one can start the analysis before an installed base exists. Suppose again that early adopters prefer product A and later adopters will prefer B; or that A is cheaper to produce early on and B is cheaper later. If an early advantage inefficiently outweighs a later advantage, we call this "*ex ante* excess inertia." One might expect that *ex post* excess inertia will tend to produce *ex ante* excess inertia, because early adopters do not sufficiently fear that later generations will strand them; but the links between the two have not been fully explored.[113]

[109] Here we assume that the network effects are real and take as given the incompatibility.

[110] Most of the literature contrasts "inertia," or the tendency to stick to an established standard or one with an early advantage, versus "momentum," or the tendency to move away from such a standard. In physics, inertia and momentum are the same thing, so some writers prefer to contrast excess inertia with "insufficient friction."

[111] Group 1 matters for welfare but makes no choices. We assume here, like much of the literature, that group 1 will not incur the switching costs required to adopt B if others do so, but will be stranded.

[112] This argument essentially comes from Farrell and Saloner (1986a) who found that ex post excess inertia and excess momentum are both possible. See also Shy (1996) and Witt (1997).

[113] Analysis of *ex ante* inertia could subsume analysis of *ex post* inertia by taking the early adopters' preference for the old product to be very strong. However, in the widely used two-generation models with each generation optimizing, this then

Thus, in Figure 2, suppose that group 1 prefers "all adopt A" while group 2 prefers "all adopt B." With strong network effects, if group 1 moves first, it gets its way – even if group 2's preference is stronger (assuming that there are no channels for efficient side payments). Moving first gives the commitment edge, and when expectations are king, commitment is a kingmaker. Thus *ex ante* excess inertia is likely at the adopters level: early adopters are pivotal and have disproportionate power. One can call this simple lock-in.

Can later adopters somehow commit to their preferred choice? Possibly, especially if network effects are not strong. For example, if group 2 is large enough to exhaust network benefits, or just does not value them highly, it may not follow group 1's choice; and so a smaller group 1 may adopt B so as to be compatible with the anticipated choice of group 2. In this case, therefore, group 2 is pivotal.[114] But such "reverse lock-in" is rarer and weaker than simple lock-in. Thus Farrell and Saloner (1985) found that, given all players' preferences, each player is better off moving earlier. This "New Hampshire Theorem" strongly suggests *ex ante* excess inertia.[115] Choi (1994) and Choi and Thum (1998) show that pre-emption competition for the New Hampshire first-mover advantage can make adoption inefficiently fast (when it is socially desirable to wait, e.g. to learn competing standards' relative technical quality).

Starker results emerge if early adopters do not care about compatibility with later adopters, so that reverse lock-in does not arise. Arthur (1989) studied sequential adoption when each adopter's payoff reflects his idiosyncratic preferences between A and B, but also increases in the number of *previous* adoptions of the technology he picks. A technology favored by (enough, but perhaps not very many) early adopters can then readily become permanently "locked in." Intuitively, if the relative network sizes ever become lopsided enough to outweigh the strongest idiosyncratic preferences, that is (in Markov chain language) an absorbing state. This model would suggest that long-run network technology choice is determined by a first-mover advantage and by "historical small events."[116] In this interpretation, as in the "butterfly effect" in nonlinear dynamics,[117] strong network effects make technology adoption largely random.

In summary, among adopters, early movers have the pivotal strategic advantage: there is a reasonable (though not conclusive) presumption of *ex ante* excess inertia. As we will see in 3.7 below, however, this need not imply that early advantages confer sustained success when "sponsors" of competing standards engage in penetration pricing.

3.5.3 Positive Feedback and Tipping

We have seen how early choices are powerful, able either to help coordination or to wield disproportionate influence. Thus when there is little functional difference between two network products, any early lead in

assumes away any coordination problems once the installed base exists, and hence (see Farrell and Saloner 1986b) predicts ex post excess inertia because the only externality is the stranding effect. This appears to be the force behind the results of DeBijl and Goyal (1995).

[114] Holmes (1996) shows how adopters who care less than others about network effects (but have similar preferences between products, or in his case locations) can lead a transition, and argues that this may have been important in the migration of the US cotton textile industry. Large groups that can successfully coordinate internally may well thus be prime candidates to move early, and perhaps therefore to be pivotal movers and get the best deals. (See Bresnahan's (2001) discussion of AOL's adoption of Internet Explorer.)

[115] However, *ex ante* excess momentum is also possible, because the outcome depends only on ordinal preferences and not on their intensity.

Farrell and Saloner called this the "New Hampshire theorem" because political commentators note that holding an early primary, as New Hampshire does, enhances a state's effective importance when bandwagon effects are important in a national election.

[116] In this spirit, David (AER 1985) suggested that the QWERTY typewriter keyboard dominates because of some highly contingent and seemingly small historical events, and not because of efficiency. Liebowitz and Margolis dispute the account, in particular the claims that the Dvorak keyboard is plainly more efficient. Shapiro and Varian (1999) argue that while network effects can sustain an inferior product, the network effects in typewriter keyboards are dissipating.

[117] With highly nonlinear dynamics, a butterfly flapping its wings might cause a hurricane years later and thousands of miles away.

adoptions (whether strategic as in section 3.4.2 or accidental as in 3.5.2) will tend to expand rather than to dissipate. This is often expressed by saying that network markets are "tippy": one network will tend to acquire an overwhelming market share, and early small leads may determine which wins.

For instance, consider identical players who will adopt one or another network product. If there is a continuum of adopters who do not inherently care which product they adopt, and only want to coordinate, then there are three static pure-strategy Nash equilibria: all adopt A, all adopt B, and an equilibrium in which half adopt A and half adopt B (and all are indifferent).

To go beyond the multiple equilibria, consider the following dynamics: market shares are randomly perturbed, and at each instant some adopters can change their move in response to current shares. Then as soon as the shares depart from equality, all those who can choose will adopt the majority product.[118] Under these dynamics at least, the equilibrium is highly unstable. Such instability persists even with some horizontal product differentiation, as Arthur's (1989) model suggests.

Moreover, when there are a finite number $K = 2k$ of adopters rather than a continuum, the same force prevents it being an equilibrium at all. For any single adopter of (in the proposed equilibrium, say) A must be joining $k - 1$ other adopters of A against k adopters of B: why not join the B group and get the benefits of being part of a group of $k + 1$ instead? This would apply to an adopter in the middle of the preference distribution if there were a symmetric divided equilibrium. Thus a static formulation of tippiness is that the interior pure-strategy equilibrium does not even exist.[119]

3.6 Monopoly Pricing

Having discussed the demand side of network markets – that is, adopters' (buyers') behavior given the offers made to them – we turn next to the supply side. We begin with monopoly pricing, with a single price and with price discrimination. Then we discuss how a seller might use contingent contracts to solve the coordination problem, whether of a single network good or of one that competes with others.

3.6.1 Single Monopoly Price

Consider a monopolist setting a single price for a single network good. Following Leibenstein (1950), we phrase this discussion in terms of a "demand curve," putting much of the difficulties of adopter behavior into that "curve." It is easiest to think of the number of groups, K, as large.[120]

To begin with, separate actual and expected network size. Let $w(x, x^e)$ be the marginal buyer's gross willingness to pay when sales are x and each buyer expects network size x^e. This w function contains a lot of demand information, but does not determine the demand curve: we must also specify how buyers form expectations. If each buyer takes as given expected network size x^e, demand is $p = w(x, x^e)$. If expectations are fixed, this is the appropriate demand curve.[121] But if buyers' expectations respond to price, then the seller does not face a fixed value of x^e.[122]

If, given x^e and p, exactly $x = x^e$ buyers would adopt, then x^e is a rational-expectations network size at price p. Because (as we saw above) there can be multiple rational-expectations equilibrium values of x at a given price, the rational-expectations demand curve $p = w(x, x)$ can have upward-sloping portions, even though each fixed-expectations demand curve $p = w(x, x^e)$ is downward-sloping as usual. This logic

[118] This assumes that they do not have perverse beliefs about how shares will evolve.

[119] See e.g. Katz and Shapiro (1985, 1994).

[120] For more on this formulation see e.g. Rohlfs (1974, 2001), Katz and Shapiro (1985), or Economides (1996).

[121] In particular, to calculate surplus as the integral under a demand curve, one should use the demand curve holding x^e fixed at its actual value.

[122] This assumes that buyers observe these choices before forming their expectations; delicate timing issues might arise, and the commitment to scale might take place by means other than simply selling a lot.

lets the seller in effect "affect expectations," x^e, by setting a price such that x^e is an adoption equilibrium at that price.[123]

Because equilibria with more adoption are Pareto preferred, some argue (in the case of a single network good) for selecting the most-adoption branch of the demand curve: essentially trusting that the coordination problems are reliably solved.[124] We doubt that coordination is always so easy. But however one "selects" the equilibrium, there may well be discontinuous changes in behavior as a parameter such as cost varies continuously: network effects can cause discontinuous dynamics, as in catastrophe theory.[125] A product subject to network effects may become cheaper or better over time in a continuous way, yet may rather suddenly acquire "critical mass" and "take off."

The rational-expectations demand curve is more elastic at any point x than the underlying fixed-expectations demand curve with $x^e = x$, because an increase in x^e raises each adopter's willingness to pay. For the same reason, a monopolist somewhat internalizes the network effects. This argument is clearest if the monopolist can price discriminate. Segal (1999) shows that a perfectly discriminating network monopolist who can commit to a pattern of offers will implement the first-best outcome. Intuitively, a monopolist who captures all the surplus will want to maximize that surplus. But this relies on commitment: the monopolist must convince buyers that it will sell to all who should efficiently adopt, typically including some unwilling to pay as much as marginal cost.[126]

Nevertheless, a monopolist setting a single price will, as for a conventional good, always set it above marginal cost, so the single-price monopoly output cannot exceed competitive output.[127] Indeed, the standard (single-price) monopoly deadweight loss may be more severe when network effects apply.[128] Not only are marginal adopters deterred from adoption, but also the surplus created by inframarginal adoption is reduced by the quantity reduction: with network effects, quantity is a dimension of quality.

3.6.2 Pricing to Complementary Customers

With network effects, each user's adoption complements others' adoption. Recognizing this, a seller should take into account in pricing to each class of users not only the user's own demand elasticity but also the cross-effect of the user's adoption on others' demand.

Such pricing to complementary (groups of) users arises in two ways in network economics. The first interpretation is that these groups are similarly situated or "peers," such as the groups in section 3.2: consider a seller pricing a network good to "early" and "late" buyers. The second interpretation is pricing to two sides of a market with indirect network effects, as with credit-card users and merchants. In either

[123] The term comes from Economides and Himmelberg (1995), although the construction dates back to Leibenstein (1950).

[124] See for instance Katz and Shapiro (1986a,b) within generations; Economides (1996) in a static model. Of course (assuming price is at least incremental cost) this is closely related to letting the seller select among multiple adoption equilibria, as studied by Segal (2001).

[125] Indeed, as in catastrophe theory, there may well be no everywhere-continuous selection of adoption equilibrium, even if there is everywhere a locally continuous selection. See Farrell and Shapiro (1992), Economides and Himmelberg (1995), Gandal (2000) for possible examples of sudden success of network products potentially explained by this phenomenon. Liebowitz and Margolis (199x) note some caveats in making such an interpretation.

[126] Segal describes this as a problem of *observation* by one buyer of simultaneous offers to others ("private offers"). Other literature has focused on *time-consistency* (the monopolist makes offers over time and there are marginal adopters in the future). Liebowitz and Margolis (JEP 1994) earlier found a special case of Segal's result: if buyers are identical, then a monopolist will sell to all buyers (if adoption is efficient): issues of discrimination and commitment vanish in this case. Panzar and Wildman (1995) also noted that a discriminating network monopoly may voluntarily subsidize very price-sensitive adopters.

[127] Economides and Himmelberg (1995).

[128] See Farrell and Shapiro (1992) for a linear example; but see also Lambertini and Orsini (Southern EJ 2001).

interpretation, sales to each group raise the other group's demand.[129] [130]

Not surprisingly, a group will get a low price if its own demand elasticity is high, and if the other group's demand responds very positively to this group's adoption choice. If both the own (price) and the cross (quantity) elasticity effects are large, the low price could easily be below cost.[131] [132]

3.6.3 Monopoly Penetration Pricing

Just as in switching-cost markets, *penetration pricing* is important. The seller sets a low price to early buyers, and then can charge more for a more valuable product when later buyers arrive if the product has taken off. The previous paragraphs suggest exploring *monopoly* penetration pricing based on the complementarity between early and late adopters (later, we will examine competitive penetration pricing).

If a monopoly can commit to price paths, there is a rational-expectations argument that it should not generically engage in penetration pricing. Early buyers should be just as willing to pay for a product whose network they know will quickly grow, as later buyers are willing to pay for one that is already big.[133] Thus it might seem that if monopoly penetration pricing is to be successful, it becomes unnecessary. However, if buyers are pessimistic (or perhaps just uncertain), they may demand to see actual adoptions before they will pay an amount reflecting the long-run network value of the product.

If the monopolist cannot commit to future prices, then its first-period price affects first-period demand directly and later demand indirectly (through installed base). As a result, the monopolist has an incentive to set a lower first-period price.[134] This is rather like monopoly with switching (setup) costs.

Reverse penetration pricing Alternatively, a seller might optimally (with full commitment) set a high early price and low later price, if the cheapest way to enlist critical mass is to cater to later adopters: reverse penetration pricing. But such prices may not be credible: once early adopters have bought, expecting low later prices (and thus widespread later adoption), the seller may not *ex post* want to lower prices.[135]

A monopoly might then find it profitable to invite (delayed, and/or imperfect) competition.[136] Indeed, firms hoping to establish a *de facto* standard often try to recruit partners into an alliance.[137] While

[129] See e.g. Schmalensee (2001) and Rochet and Tirole (2000) and Caillaud and Jullien (2001). For other examples, consider auction houses (traditional or eBay) setting buyer and seller commissions, or matchmakers pricing to men and women. Adobe sets pricing to writers and readers of PDF files. Newspapers, broadcasters, and web sites set prices to readers and to advertisers. Rosse (1967) argued in early work that the newspaper industry is subject to strong network effects because of the positive feedback between the size of readership and the demand on the part of advertisers.

[130] It is convenient to think of the seller as setting quantities rather than prices. Suppose that the marginal member of group i is willing to pay $w_i(x_1, x_2)$ if sales are (x_1, x_2). When the groups are peers, w_i may depend on aggregate adoption, most simply on $x_1 + x_2$; on the other hand, if the groups are two sides of a market, w_1 may be increasing in x_2 and decreasing in x_1. Thus in the most direct sense each seller loses when other sellers arrive at the market. There are indirect network effects if he gains more than that loss when we account for the buyer response to the presence of more sellers. In either case, simplifying by assuming zero marginal costs, the seller sets (x_1, x_2) to maximize $x_1 w_1(x_1, x_2) + x_2 w_2(x_1, x_2)$.

[131] Such a price below cost need not suggest predation (indeed there can be no predatory motive in this model). On predation with dynamic increasing returns, see e.g. Cabral and Riordan (1997) and Farrell and Katz (2001).

[132] Every MBA student knows to "give away the razors and sell the blades," but few can intelligently discuss why it is that way around and not the other. [ARE THESE NETWORK EFFECTS, SWITCHING COSTS, OR JUST COMPLEMENTARITIES?]

[133] This is analogous to the familiar point in learning-by-doing models (where costs are a function of cumulative experience) that the appropriate concept of "marginal cost" involves foresight, so one should not expect prices to decline over time merely because naively calculated marginal costs are falling.

[134] Cabral, Salant, and Woroch (1998) study the theory of monopoly penetration pricing of durable network goods when buyers have rational expectations. In certain classes of example, they find that Coase-conjecture price dynamics tend to predominate over penetration pricing: that is, prices fall rather than rise over time, especially when there is complete information. Bensaid and Lesne (IJIO 1996) find however that strong network effects remove the time-consistency Coase problem and cause optimal prices to increase over time. See also Mason (EER 2000) and Choi (JIE 1994).

[135] Segal (1999, 2001) argues that the ability to commit greatly affects the efficiency of monopoly provision of a single network good. See also our discussion of Katz and Shapiro (1986a) in 3.7 below.

[136] See Farrell and Gallini (1988) and Economides (Eur JPE 1996). For a somewhat different rationale, see Choi and Thum (EER 1998).

[137] Cusamano et al. (1992), Farrell and Shapiro (1993), Axelrod et al. (1995), and Shapiro and Varian (1999), discuss aspects

coalitions are important (e.g. Shapiro and Varian 1999), there has been relatively little research on the theory of nontrivial standards coalitions; most of the literature makes a simplifying duopoly assumption.[138] Commitment not to gouge adopters later (or later adopters) is also an economic rationale for the practice in formal standardization of requiring early disclosure and "reasonable" licensing of essential intellectual property.[139]

3.6.4 Contingent Contracts

As Dybvig and Spatt (1983) noted, government or a seller may be able to use contingent pricing to overcome buyers' coordination problem. Suppose the seller offers buyers a contract: "The price is $p < u(N)$ if all other buyers also adopt (which I expect); if not, the price is $p' < u(n_i)$." It is a dominant strategy for each buyer to accept this contract, whatever he may believe about other buyers' behavior. Of course, p' may have to be (perhaps far) below cost, so the seller will make a loss if buyers behave irrationally. But success depends only on buyers' individual rationality, not on their attaining the efficient (or indeed any) equilibrium.

A similar contingent contract can, in theory, similarly attract all buyers away from a Pareto-inferior coordinated equilibrium.[140] Thus, suppose that buyers expect one another to adopt B, giving each surplus $u_B(N)$, and that network effects are strong so that this exceeds $u_A(n_i)$ although it is less than $u_A(N)$.[141] A seller of A offers buyers the contract: "If x of you buy A, the price will be $u_A(x) - u_B(N) - k$." For $k > 0$, it is a dominant strategy for each buyer to accept, and the contract is profitable if all buyers do so and k is small enough. Indeed, Segal (1999) shows that, because adoption of A imposes a negative externality on those who continue to buy B, there may be excessive adoption of A in some circumstances.

It is no surprise that flexible contracting can in theory solve the coordination problem.[142] At the level of *cooperative* game theory, network effects are like ordinary economies of scale.[143] Since simple bilateral offers ("here's the price; accept or reject") enable efficient competition with economies of scale when preferences are symmetric, there must be contracts that do so with network effects; we have described one such contract.[144] The question is whether this truly solves the problem of sponsors of superior offers who will otherwise be excluded by coordination problems. While the literature has not yet fully addressed this question, it seems that contracts with such contingent pricing are not widespread, although penetration pricing gives early adopters some analogous assurance.[145]

---

of the alliance formation process.

[138] See however Cremer et al. (2000), Farrell and Shapiro (1993), Economides and Flyer (1995), Farrell Monroe and Saloner (1998), Axelrod et al. (1995), Gandal and Shy (2001).

[139] See e.g. Shapiro and Varian (1999).

[140] This assumes that the incumbent network is not also using this kind of contract. Jullien (2001) considers competition between such contracts **(is this right?)**.

[141] Assume for simplicity that costs are zero and that the price of B is non-negative.

[142] This is explored further by Segal (1999, 2001) and Jullien (2001). Thum (1994) also considers how contract form affects efficiency.

[143] Each case is characterized by the condition that a coalition consisting of a seller and x buyers achieves more surplus per buyer as x increases: in cooperative game theory, this is all we specify.

So we should see a similar "risk of failure" when the economies are on the production side: and we do. The corresponding risk is that a seller who prices assuming that all buyers will accept its attractive offer will lose if for some reason many of them do not.

[144] Innes and Sexton (1994) discuss difficulties in such contracts; see also Haruvy and Prasad (J Evol Econ 2001). Segal (1999, 2001) investigates the effect of allowing a seller to select among multiple adoption equilibria.

[145] One problem with such a guarantee contract, if adoption takes place over time, may be going back to early adopters to top up their payments. If that is too hard, and if promising refunds for inadequate network growth is also problematic, Rohlfs (2001) notes that one imperfect solution is usage-based pricing. That is, if each adopter's use of a telecommunications product, say, is proportional to the value he derives from it, then traffic-sensitive pricing may solve the chicken-and-egg problem even at the cost of inefficiently deterring usage given network size.

3.7 Competitive pricing with network effects

When competing products are incompatible, our discussion of buyer behavior above suggests that firms will vie to control expectations. In general, competition will largely focus on pivotal customers; because of adopter-level inertia, this will very often mean early adopters – much as with switching-costs, where competition is largely for early purchases. Central questions are how reliably more efficient firms win and whether profits reflect only their efficiency advantage. These issues are importantly dynamic, but we first see what can be gleaned from a simpler static formulation.

Sometimes firms may control expectations through offering a better deal (expectations "track surplus"), and in a static model this defuses competitive concerns about network effects. Dynamic competition when expectations track surplus is a lot like competition with switching costs.[146] But while cheap talk and sequential adoption will help make expectations track surplus, it seems over-optimistic to assume that better deals will always attract expectations. And when preferences are not "similar," whether because of sheer taste differences, sunk investments, or discriminatory prices, neither cheap talk nor sequential adoption can as effectively help adopters coordinate. Thus it is important to understand how competition works when expectations do not necessarily track surplus.

In 3.7.1 we note how incompatibility affects incentives to win one customer when contemplating competing for another. In 3.7.2 we discuss static competition with vertical (or no) differentiation and three ad hoc but useful patterns of expectations. In 3.7.3 we discuss dynamic competition with (at each date) vertical differentiation. In 3.7.4 we discuss static competition with horizontal differentiation. Finally, in 3.7.5 we discuss dynamic competition with horizontal differentiation.

3.7.1 Incentives to win a customer

Under compatible competition if a firm wins a customer from its rival it does not directly affect competition for another customer: both firms are in any case offering the same network size. If it wins a customer who would otherwise not have bought at all, it improves the network-size "quality" of its own offering but also of its rival's. Thus network size is a public good in compatible competition: see Kristiansen and Thum (1997).

In contrast, in incompatible competition, if a firm wins a customer who would otherwise not have bought, it improves the quality of its own offering and does not affect its rival's (this is the internalization of network size discussed above in the monopoly context). On the other hand, if a firm wins a customer from its rival, it strengthens its own offering *and* weakens its rival's in competing for other customers, so that there is an anticompetitive (rival-weakening) incentive: see Farrell and Katz (2001).

3.7.2 Static competition with vertical differentiation

A simple case of incompatible competition is that of no differentiation or purely vertical differentiation: either a cost difference or a quality difference valued at the same amount by all consumers.[147] For this case, Farrell and Katz (1998) studied the implications for price and quality competition of three simple expectations patterns: expectations may track surplus, track quality, or stubbornly favor one firm.

Expectations may *track surplus*: each buyer expects all others to buy the product that, network effects held constant, offers the greatest surplus. Price competition then works just as it would if the products were compatible. The efficient product wins, and consumers get the same surplus as they would if the

[146] As with switching costs, the dynamics come about through installed bases, and the reason those matter is that people in the installed base have made sunk investments and won't lightly switch.

[147] Baake and Boom (), and Bental and Spiegel (1995) discuss static models of competition with network effects and quality differentiation when consumers' willingness to pay for the improvement varies.

second most efficient product were offered at cost and adopted by all. Consumers capture the network effect, while the best product wins and the winning seller captures its efficiency advantage. Quality competition also is therefore just as under compatibility.

Sequential adoption often helps expectations track surplus. Thus suppose that firms set prices just once and then there is a sequence of adoption choices by small cohorts. If (given prices) adopters have similar preferences (agree on which product offers them more surplus if all adopt it), then the unique subgame-perfect equilibrium is adoption of that product, as we saw above.

But this changes dramatically if sponsors can strategically adjust prices to later adopters *in response* to earlier adoption choices. Suppose for instance that product A has higher quality (and equal costs), and that the higher quality outweighs the network gain from adoption by a single cohort, though it is outweighed by the network effects overall. Then, A will not fail just because a *few* buyers adopt B instead. Rather, such a loss will lead A's sponsor to improve its offer to subsequent adopters: it can do what it takes to win.[148]

Reasoning in that way, each adopter will recognize that even if he and his cohort adopt product B, product A will still win the rest of the market. Thus no buyer is pivotal, so the price terms offered any buyer (or cohort) should not affect expectations. So rational expectations will *track quality* and ignore prices.[149]

Then, of course, price competition is muted. Specifically, if A has higher quality and therefore attracts expectations for all price combinations for which that is an equilibrium for adopters, then A can win current sales if $u_A(N) - p_A \geq u_B(n) - c_B$, where we assume for convenience that each group or adopter is of size n and also make the standard Bertrand-style assumption that a losing seller is willing to price down to cost. We can rewrite this as $p - c_A \leq [u_A(N) - u_B(N)] + [u_B(N) - u_B(n)] - [c_A - c_B]$. So the more efficient firm's reward is the sum of its actual (cost and/or quality) advantage and the network effect. If A *can* make consumers a significantly better offer than can B, then A need not *actually* make nearly as good an offer as B would be willing to! Consumers end up with *less* surplus than they would get from all adopting the losing technology B priced at cost.[150]

Of course, when such lucrative expectations track quality, firms will compete intensely on quality. Consumers gain the benefit of the additional quality effort incurred by the second most successful (at quality) firm. The network effect accrues to the winner, and/or is dissipated in quality competition, which can therefore be socially excessive.

Quality, however, is only one force that might make consumers expect a product to be able to win (almost all) the market even after (out of equilibrium) losing a round or two – making expectations stubborn and unresponsive to price or performance.[151] For instance, the same dynamic logic as above might suggest this outcome if one firm plainly *could* dramatically improve its product if necessary — even if it never actually does so. Other forces might include deep pockets, history or reputation, a convincing road-map for future products, control of a key complement, control of formal standards efforts, or marketing activity. A seller favored by expectations can extract profits commensurate with the network effects, and may thus be able profitably to control the market even with a substantially inferior product or offering — provided, crucially, that such inferiority does not cause it to lose control of expectations.

When expectations stubbornly favor one firm, it has monopoly-like incentives for quality improvement,

[148] It follows that the inferior product B will not engage in penetration pricing: there is no follow-on gain to winning a cohort or two. This analysis is essentially taken from Katz and Shapiro (1992).

[149] More generally, expectations will focus on firms who *can* make better offers, not necessarily on those who *do* so.

[150] This is an instance of the principle that pivotal adopters get the surplus: when there are no such buyers, firms can keep the surplus. (But see Raskovich 2001 for a bargaining argument to the contrary.)

[151] There is of course a limit to stubbornness: in particular, if it is not a (static) equilibrium for adopters to adopt A, because A's price is so high, then it is irrelevant what each buyer expects others to do.

while its rivals can profit from improving their quality only if they can innovate so dramatically as to overturn the expectations, giving them potentially very strong incentives for dramatic innovation but no incentive for ordinary innovation.

3.7.3 Dynamic competitive pricing with changing vertical differentiation

As we saw above, early adopters are pivotal when network effects are strong. If early efficiency advantages, such as an initial installed base, determine offers to these pivotal early adopters, then a technology with an early lead will beat a technology with a bigger potential advantage later, provided the network effect is stronger still. This is the New Hampshire Theorem's version of excess inertia.

But what if "sponsors" of nascent challenging technologies compete for early adopters through penetration pricing? Will such competitive penetration pricing yield efficient adoption choices? Clearly, it can do so only if there is penetration pricing, so we assume first that competing technologies are all sponsored.

Thus suppose that A has costs a_t in period t, while B has costs b_t, and that network effects are strong: second-period adopters would follow first-period adopters if both products were priced at cost, and will pay r for a product compatible with first-period adoption. Finally, suppose that if a firm fails to win first-period sales, it exits (it knows it will lose in the second period).[152] Then A will be willing to price down as far as $a_1 - (r - a_2)$ in order to win first-period sales, while B is willing to price down to $b_1 - (r - b_2)$. Consequently, second-period efficiencies feed through into first-period penetration pricing, and the firm that can more efficiently provide the good in both periods wins sales in both periods, if each cohort optimally coordinates internally and first-period buyers correctly foresee second-period behavior. In this central model there is neither ex ante excess inertia nor momentum.[153] That is, the more efficient provider wins the market, and the pivotal (first-period) adopters get the benefit of competition.

Second-period efficiency can feed through *more* strongly than this into first-period penetration pricing. In Katz and Shapiro (1986a), a first-period loser does not exit but continues to constrain pricing. Thus the second-period prize for which A is willing to price below its cost in the first period is $b_2 - a_2 + \beta$, where β represents a network-size advantage;[154] similarly B is eyeing a second-period prize of $a_2 - b_2 + \beta$. So firm A wins first-period (and hence all) sales if and only if $a_1 - [b_2 - a_2 + \beta] \leq b_1 - [a_2 - b_2 + \beta]$. Second-period efficiency is double-counted relative to first-period efficiency, leading to *ex ante* excess momentum.[155] That is, the more efficient provider does not necessarily win the market: there is a bias in favor of a technology that is cheaper in the second period.

But there is ex ante excess inertia if feed-through is weaker than in the central model. There is *no* feed-through when the standards are unsponsored (so that firms cannot later capture gains from establishing a better product). Less extreme cases also seem plausible.[156] Katz and Shapiro (1986a) also noted that

[152] Farrell and Katz (2001).

[153] Welfare may still be lower than under compatibility, although firms have an incentive to achieve compatibility in that case. See Katz and Shapiro (1986b). The network effects accrue to consumers – but not uniformly: as we have seen before, it is the pivotal buyers who get the surplus.

[154] Specifically, β is the difference in value between a network of all consumers and one consisting only of second-generation consumers. With strong network effects, β exceeds second-period cost differences. Shapiro (GeoMasonLawRev 1999) argues that even if group 2 is quite large, β may remain large if group 1 is also large, and that this is a difference from the case of production-side economies of scale.

[155] This is why Katz and Shapiro (1986a) find excess momentum (they call it the "new-firm bias") with sponsored products when network effects are strong. When network effects are weaker, they found a new-firm bias for a different reason: a commitment problem, as in Segal (1999) (see 3.6 above). The firm with the second-period disadvantage may not win second-period sales even if it wins first-period sales; but the ("new") firm with the second-period advantage certainly would. The two firms are then offering different prospective network sizes to first-period customers. The "old" firm would like to be able to commit to winning the second period after a first-period win, but cannot do so.

[156] For example, with downward-sloping demand and exit, second-period customers are better off if the firm with a second-period cost advantage wins the first period: this efficiency advantage does not fully enter second-period profits and so will not feed through. Feed-through will also be weakened (as in switching-cost models) if firms cannot lower first-period prices enough to pass through all prospective ex post profits to the pivotal early adopters (e.g. because of borrowing constraints, or

asymmetric feed-through – as when one product is sponsored but another is not – will bias the outcome toward the sponsored product.

Thus (once again) when prices are exogenous, network effects can make the preferences of pivotal adopters over-weighted relative to non-pivotal adopters'; there is a presumption that early adopters are pivotal, and hence that the dynamics of adoption with given prices favor products with early advantages or installed-base leads. When prices are endogenous, as with penetration pricing, efficiencies in serving non-pivotal adopters can feed through into willingness to price low to pivotal adopters. This feed-through can be zero (as with unsponsored products), correct (as in the central model above), or excessive (as in Katz and Shapiro 1986a and Jullien 2001), or presumably in between.

While the theory thus is agnostic, we suspect that inadequate feed-through is a more robust possibility than excessive: spillovers that prevent a firm from capturing the whole benefit of its penetration pricing are probably more widespread than the effects identified by Katz and Shapiro. Moreover, if the first-period efficiency advantage is not cost efficiency but an installed base, and if first-period buyers' expectations track past success (the installed base) rather than optimally coordinating, then the incumbent product will win even if the entrant is modestly more efficient overall.[157]

Incompatible entry even by a sponsored product is often hard, but not clearly *too* hard, *given* the incumbent's installed base and *given* incompatibility; however, for the reasons above, it may well be too hard.

A Switching-Cost Analogy The models above have close switching-cost analogies, although the switching-cost literature has not focused extensively on efficiency differences between firms. With costs as described above and no network effects but a switching cost s, suppose first that each buyer expects to face a second-period price p_2 that is independent of which seller he is locked into. Then of course he will buy the lower-priced product in the first period. If the buyer's belief about second-period pricing is right (for instance, if the buyer's reservation price r is low enough that switching can never pay, so $p_2 = r$), then seller A is willing to price down to $a_1 - [p_2 - a_2]$ in the first period, and similarly for B. Hence, the firm that has lower life-cycle costs makes the sale. This is the switching-cost analogy to the central model above.[158]

But if second-period prices are in fact constrained by the buyer's option to switch, then A will price at $b_2 + s$ in the second period if it wins the first, while B will price at $a_2 + s$ if it does. If "myopic" buyers do not foresee this difference then second-period costs are double-counted relative to first-period costs: this is an asymmetric version of the model in 2.2.1 above, and is the switching-cost analogy to Katz and Shapiro (1986a). Finally, if second-period prices are constrained by the option to switch and buyers have rational expectations (they know what constrains prices and know the firms' second-period costs), then the buyer will buy from A only if its first-period price is at least $b_2 - a_2$ lower than B's, and again the firm with lower lifecycle costs wins.

Investment Incentives Incentives to reduce costs in some of these models seem complex, since each firm's second-period cost affects the second-period price that the other would be able to charge.

because negative prices attract worthless demand).

[157] The rational expectation and optimal coordination assumptions are relaxed in Farrell and Katz (2001). One could interpret this as a loose reduced form of modeling many cohorts: compare our discussion in 3.7.2 above following Katz and Shapiro (1992).

[158] See also section 3.2 of Klemperer (1995).

3.7.4 Static competition with horizontal differentiation

We saw in 3.5.3 above that if competing network products are undifferentiated, or purely vertically differentiated (all consumers agree on the difference in value), then the market will tip: in any pure-strategy equilibrium, one network will get all the sales. With horizontal differentiation, market-sharing is of course an equilibrium if network effects are weaker than differentiation, but we still get tipping if network effects are strong. Then competition is not really "for" the business of individual consumers but is really "for the market" as a whole.

This particularly changes the nature of competition when there is some horizontal differentiation but network effects are strong.[159] With compatibility, one would then expect three conventional lessons. First, both products coexist in the market. Second, prices reflect each firm's market power due to differentiation. Third, if the seller of either product modestly improves his product, it yields modestly higher gross profits.

With incompatibility, all these lessons change. Buyers (want to) act as a group, and horizontal differentiation is overwhelmed. One product will make all the sales (if buyers succeed in coordinating, whether or not on a good offer), and at equilibrium the winning product charges all it can without losing all the sales – implying that the losing product could make all the sales by marginally cutting its price. Hence, as in undifferentiated Bertrand competition, the loser's price is presumably set at its cost.[160] And the returns to product improvement depend critically (as in 3.7.2 above) on how expectations will respond.

Thus, strong proprietary *network effects can sharpen price competition.* The network effects can make sellers bundle activities in which they have a disadvantage with activities in which they have an advantage, which can sharpen price competition. This is essentially the Einhorn (1992) point: see section 2.xx above.[161]

For example, if expectations track surplus as assessed by (say) the median buyer,[162] then strong proprietary network effects sharpen price competition. Absent network effects – or with compatible products – horizontal differentiation would soften price competition in the usual way.[163] With strong proprietary network effects, competition is as if all buyers had the tastes of the median buyer: sellers must bundle selling to distant consumers with selling to nearby consumers.[164]

Similarly, in a two-group model, as in Figure 2, suppose that group 1 prefers A, group 2 prefers B, production costs are equal, and network effects are strong. Then under compatibility, if sellers set different prices to the two groups, each seller sells to its matched group and makes profit equal to its efficiency advantage in serving that group.[165] Under incompatibility, absent coordination failure, the market will tip. While it is unclear in a static framework how buyers jointly choose which product to all adopt, neither firm has an overall efficiency advantage as big as its profits (let alone industry profits) under compatibility, so if buyers are good at coordinating then total profits seem very likely to be less than under compatibility.

When differentiation is stronger, or network effects weaker, there need not be complete tipping: niche

[159] Recall that this means that (despite the differentiation) each buyer would rather do what everyone else does than be the only adopter of his otherwise preferred product. Of course this depends on prices; probably it is most usefully assessed when each product is priced at cost.

[160] Even if the firms have equal costs, the winner might make profits, but only if it has an expectations advantage: if expectations track price, the winner can make a profit only equal to its cost advantage.

[161] On the other hand, if network effects are less than market-wide (so that each adopter cares primarily about compatibility with a limited set of others), the network effects can create protected enclaves. Other effects?? Discuss de Palma Leruth and Regibeau; Boom; Lambertini and Orsini; Grilo, Shy and Thisse (2001), Encoua, Moreaux and Perrot (IJIO 1996); Belleflamme (IJIO 1998).

[162] That is, each buyer expects that all other buyers will buy the product preferred by the median buyer.

[163] Large buyers in oligopoly markets often negotiate discounts in return for exclusivity. One possible explanation is that a "large buyer" is really a joint purchasing agent for many differentiated purchases; exclusivity commits the buyer to ignore product differentiation and thus sharpens price competition.

[164] An analogous switching-cost model has buyers in the first period knowing nothing about their second-period tastes and having no first-period taste differences.

[165] What happens under compatibility if no price discrimination? I'm not sure whether the literature has thought about this.

minority products can survive.[166] But the strategy of selling only to closely-matching buyers is less appealing than under compatibility (or without network effects), and if network effects strengthen, or the dominant network grows, niches may become unsustainable.

3.7.5 Dynamic competition with horizontal differentiation

Again, we start from the point noted in 3.5.3 above, that positive feedback and extreme market shares (tipping) can be expected in the adoption of competing (not too differentiated) network products at given prices. In particular, one can expect tipping in competition between unsponsored network products, and Arthur (1989, 1990) and Arthur and Lane (1993) indeed find that if firms' prices are independent of their market (installed base) shares while consumers' decisions depend only on past adoptions (current shares of installed base), then one product or technology will come to dominate.[167]

But just as excess inertia at the adopter level need not imply excess inertia when firms can engage in penetration pricing, tipping at the adopter level might not imply tipping when firms can price to build or exploit market share. If one product happens to get ahead, will its sponsor raise price to exploit (and hence dissipate) that lead, or will it come to dominate the market? Arthur and Rusczczynski (1992) studied this question when firms rationally set prices in a many-period dynamic game; Hansen (1983) considered a similar model. In their stochastic duopoly model, they find that if firms have high discount rates, a firm with a large market share tends to lose it by pricing high to exploit that market share for near-term profit. But if firms have lower discount rates, a firm with a larger market share continues to set low "penetration" prices to reinforce its dominant position.[168]

Thus tipping or unstable (positive feedback) dynamics is not only a feature of pure adopter behavior (given prices) but can readily extend to the case where firms strategically set prices. This contrasts with the case of switching costs, where firms with large installed bases have strong incentives to price high and where market dynamics thus tend to be stable, as we discussed in 2.3.3 above.

3.8 Endogenous Network Effects: Choosing How to Compete

Incompatibility may be inevitable, but very often it is a choice, and the firms could potentially offer compatible products instead. Then, why would a firm prefer one form of competition over another, and who gets to choose?

Compatible competition is basically familiar (although see 3.7.1 above): it rewards sellers for low costs, high quality, choice niches, etc., as well as for softening competition and excluding entrants. Incompatible competition also involves all those issues but focuses in particular on attracting complementors and affecting expectations.

Broadly, a firm will favor a form of competition that is relatively soft, or that gives great play to the firm's strengths. Thus a basic tradeoff was identified by Katz and Shapiro (1985).[169] A move to compatibility has two effects. The *demand effect* is that consumers get more value from the larger network

[166] Beige (2001) finds relatively general conditions for this to be possible.

[167] In these models, the probability of winning a consumer is a function of prices and shares of installed base; this assumption is rationalized by horizontal differentiation.

Dosi, Ermoliev and Kaniowski (1994) can obtain market sharing as an outcome from a related model in which firms do adjust prices in response to market shares but according to an exogenously imposed non-optimal rule. (CHECK)

[168] Katz and Shapiro (1992) also study subgame-perfect duopoly pricing in a many-period with network effects; they allow for foresighted consumers. But their model has no horizontal differentiation and cannot readily thus explore the dynamic response to small changes in market share.[or is it just that they don't talk about this?]. They find that, when network effects are strong enough, one firm wins all the sales. One might reinterpret their model however to predict that all sales go to the predicted winner and that random departures from equilibrium market shares are ignored unless they flip the identity of that winner.

[169] Cremer, Rey and Tirole (JIE 2000) apply the same logic to internet backbone connectivity. Mortimer (2001) also explores the quality and investment incentives generated.

that results from combining two previously incompatible networks, so total demand shifts upwards with compatibility; firms will in general capture at least part of this demand shift. But the *leveling effect* is that consumers will get the same network benefits by buying from either firm, so installed bases and expectations are no longer a competitive advantage, and competition may be more symmetric. A firm with a large installed base, or that would have a big expectations advantage, will thus tend to prefer incompatibility.[170] A firm might most prefer incompatibility when (as a practical matter) any competition for expectations is all but over and it controls expectations.

However, if expectations will track surplus in incompatible competition, no such expectations advantage arises.

Another effect arises in vertical mix-and-match models, as in 2.x above. A firm that is good at one activity but bad at another will want to be able to draw on outside partners for the latter, which may require compatibility. Similarly, as noted above, incompatibility can sharpen price competition in mix-and-match and related models; Matutes and Regibeau (1988) also stressed a strategic pricing effect. The efficiency benefits of mix-and-match provide an efficiency *demand or cost effect* that parallels the demand effect above; at the same time, sharing the supply of a complement is a *leveling* between competitors in any one component.

If firms disagree on compatibility, who chooses? A firm may be able to prevent (or charge for) compatibility by asserting intellectual property protection on important aspects of product design needed to ensure compatibility, or by changing its product design often and keeping interface information secret. In other cases, a firm that wants compatibility may be able to impose it.[171]

Sometimes it is not firms but governments that choose compatibility standards. In 3.10 below we discuss two motives for public policy on compatibility that could arise in a closed economy. Compatibility policy can also serve as trade policy.[172]

3.9 Compatibility Policy

Policy intervention to ensure or facilitate compatibility may avoid coordination failures or certain externalities. It can also strengthen competition. These two goals raise very different issues.

3.9.1 Compatibility as Coordination Policy

If adopters are inefficiently splintering (failing to standardize) because of confusion, or are adopting too slowly because of fear of coordination failures, an indicative (not compulsory) standard could in theory be beneficial; indeed, this is largely what non-governmental (and sometimes government) standards organizations provide.[173] If failure to standardize is not a coordination problem, but is due to horizontal differentiation and the inefficiency is because adopters of each product ignore the externality on adopters of the other, then a compulsory standard could be desirable; although recall that, as we noted above in 3.3.2, the externality is ambiguous and relatively weak when there are competing network goods.[174] If adoption

[170] Katz and Shapiro (1986a,b, 1992). Regibeau and Rockett (IJIO 1996) discuss how firms' incentives to choose compatibility may change over time.

[171] Biologists have analogously studied "arms races" including the competition between toxic species who would like to signal their identity and mimics who would like to pool.

[172] See Gandal and Shy (2001), Jensen and Thursby (1996), Crane (1979), Matutes and Regibeau (1996).

[173] Although many discussions of standards for mobile telephony assert that second-generation adoption in the US has been slow for these reasons, an indicative standard was adopted (by an industry association, not government).

[174] This is when there will be some adoption of competing standards in any case. The externality at the margin when there is tipping is not murky.

Some Americans argue for "English only" laws based on a network externality; however, at the same time and not so far away, Canadians intervene to discourage de facto standardization on English. See Church and King (199x).

is too fast in order to grab first-mover status, then an indicative standard might not help; a compulsory standard would (at some cost).

Converters or adapters can let adopters gain part of the network benefit from otherwise incompatible products and can thus defuse externalities and coordination problems. Perfect converters would moot the whole issue of coordination (a single-network externality issue would remain), but typically converters are imperfect and costly. Because the equilibrium choice of compatibility or incompatibility can be non-optimal in complex ways, it is not surprising that the introduction of converters, which mitigate incompatibility, involves second-best issues.[175]

3.9.2 Compatibility as Competition Policy

As we have seen, with (strong) network effects firms compete over sales to large groups of buyers rather than over single transactions. So, as with switching costs, certain naive competitive fears do not hold up: one must not jump from the fact that each buyer is constrained by the need to conform with other buyers to the conclusion that there is an overall competitive problem. Even if each buyer must follow the crowd, something drives the crowd's choice, and it can be the very factors (such as price and quality) that any one buyer would like to respond to. Thus network markets with incompatible products can be more competitive than they may look; a stark theoretical illustration is the static model where expectations track surplus. Competition for the market *can* work well.

But while proprietary network effects *need* not cause competitive problems, they probably do make competition more fragile.[176] Because incompatible entry is hard, there may be much more incentive for monopolizing strategies such as predation or merger than there is in markets in which easy entry limits any market power.[177] This is true whether or not the inertia is "excess," and even if buyers optimally coordinate subject to the limits imposed by an installed base that cannot readily shift. Competition policy in such cases should probably not aim to pick winners or engage in strategic subsidies,[178] but surely should recognize the fact that competition is difficult. A central finding in the Microsoft case was that the "applications barrier to entry" makes entry difficult but that market developments might have lowered that barrier: at least with optimal coordination, network effects are not absolute barriers.

If buyers do not optimally coordinate, problems are more severe. Entry may be still harder (if expectations favor incumbents); and competitive pressure (among existing firms) to offer the highest quality and lowest prices is weakened if expectations do not reliably track surplus. An incumbent favored by expectations may go on profitably controlling the market even after the entry of a more efficient rival (even evaluating efficiency taking as given the installed base and incompatibility). This makes the prospects for recoupment stronger than in the more conventional case of large economies of scale combined with some sunk costs.

---

[175] Thus, Farrell and Saloner (1992) found that converters can reduce static efficiency; Choi (1996) finds that converters can block the transition to a new technology. See also David and Bunn (1987), Kristiansen (1998), and Baake and Boom (2001).

[176] See for instance Anton and Yao (1995) and Lemley and McGowan (1998).Liebowitz and Margolis (Harv J Law Tech 1996) argue that the concern is generally overblown. Katz and Shapiro (1999) discuss antitrust in software markets, where network effects are often important. Economides and White (1994) stress parallels with conventional vertical antitrust questions.

[177] But while predation may be more likely in network markets, it is not so clear how to control it: because of the legitimate real intertemporal links and other complementarities, simple conventional "cost" rules against predatory pricing in a network market are unlikely to be efficient. See section 3.6.2 above, Cabral and Riordan (1997), and Farrell and Katz (2001).

One non-price predation strategy might involve (even truthful) product preannouncements: such preannouncements might close a narrow window of opportunity for an entrant, by encouraging adopters to await an updated product from the incumbent rather than go with an otherwise attractive entrant's product: Farrell and Saloner (1986a). But Fisher (1991) argues that it will be very hard to diagnose.

[178] Arthur (1988) might seem to suggest that subsidies or other interventions to change the standard are the main tool of policy, but he clearly recognizes the difficulties in such policy. Bresnahan (2001) stresses that antitrust policy in network industries need not (and should not) push for one standard over another or be based on a theory that the wrong standard has been chosen or that an old standard has lasted too long: rather, it should ensure a full market test (even though we know that test is imperfect).

Differences in preferences among consumers, which are to be expected especially when there is an installed base or other vested interests, make coordination problematic. The literature has stressed that conflict causes coordination difficulties in the context of formal standard-setting: the cheap-talk mechanism for coordination performs much less well, both in theory and in practice, when preferences are not similar. At the same time, such institutions are important and it is a sensible goal of policy to help them work well and protect against their abuse.[179]

Even if expectations will track surplus ex post, there may well be collective switching costs. Therefore, as in other switching-cost models, incompatible competition tends to feature fierce competition for pivotal adoptions and little competition for others. As we saw in 3.7.3 above, the degree of effective feed-through of prospective profits from dominance into competition for dominance is not always straightforward. More generally, competition for the market is not the same as, and can be less efficient than, competition in the market.[180]

Consequently, firms with large installed bases or who will attract expectations may favor incompatibility in order to weaken competitors who might be much more effective with compatible than with incompatible products. Competition policy may thus limit firms' ability to insist on incompatibility,[181] perhaps giving rivals rights that help them insist on compatibility (such as rights of reverse engineering) even despite intellectual property protection.[182] It may also try to prevent the emergence (via mergers) of firms with such large positions.[183]

Interface or vertical compatibility standards could help firms leverage market power from one vertical layer to another. Whether they rationally would want to do so is a complex economic question; the "one monopoly rent theorem" argues not, and in fact firms with market power in one layer often choose to encourage competition in complementary layers.[184] Our point here is simply that the leverage question may arise particularly strongly when interfaces matter, because they can help a firm exclude independent complementors/competitors *if* it chooses to do so.

4 Conclusion

Switching costs and network effects create fascinating market dynamics and strategic opportunities. They tie together trades that are not readily controlled by the same contract: future trades in the case of switching costs, and trades between the seller and other buyers in the case of network effects. We have stressed that the result *can in theory* be efficient competition for larger units of business – "competition for the

[179] The FTC brought a complaint against Dell in 199x, for instance, arguing that Dell had abused the standards process by failing to disclose intellectual property rights in a proposed standard.[check and cite precisely]

[180] Demsetz (1968) is the most popular citation for discussion of competition for the market, although the idea goes back at least to Chadwick (18xx).

[181] In theory firms may anticompetitively seek compatibility, as in Einhorn's (1992) model. This does not seem to have become a major concern of competition policy.

[182] Samuelson and Scotchmer (2001) discuss how US courts and the European Commission have been sympathetic to rights of reverse engineering for compatibility (interoperability). Lemley and McGowan (1998) discuss cases in which courts or competition authorities have allowed firms to achieve compatibility despite intellectual property or other arguments by incumbents who would have preferred incompatibility. Farrell (1989, 1995) describes some arguments for weaker intellectual property protection in network industries; Farrell and Woroch (1995), joined by thirteen other economists, argued to this effect in the Lotus-Borland case before the Supreme Court.

[183] One concern about the proposed merger of MCI Worldcom and Sprint was that the combined firm might have an incentive to deny compatibility in the internet backbone market for this reason. [??see the Complaints by EC and DOJ??] See Cremer et al (2000).

[184] The "one monopoly rent theorem" points out that a monopolist in good X shares consumers' interests in efficient and competitive provision of a strictly complementary good Y, and will often have no incentive to allow or engage in anticompetitive behavior in Y. However, Bresnahan (2001) stresses the possibility that such a monopolist (Microsoft in his analysis) may want to prevent the emergence of a dominant complementor because it could be a catalyst for future competition.

It is important to distinguish "open interfaces" meaning that a firm opens its interface to all complementors, from "open standards" meaning compatibility with competitors.

market:" neither switching costs nor network effects are inherently and necessarily problematic. But they very often make competition, perhaps especially entry, less effective. And policymakers should look particularly carefully at markets where incompatibility is strategically chosen rather than inevitable.