Network Effects and Switching Costs:
two short essays for the new New Palgrave

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

Paul Klemperer
Edgeworth Professor of Economics, Oxford University,
and
Member, UK Competition Commission*

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

First draft: 2004
This draft: March 2005

Abstract

We briefly survey the economics of network effects and switching costs (in 3,400 words). For comprehensive coverage of the same ground see Farrell and Klemperer’s 60,000-word contemporaneous survey, available at www.paulklemperer.org.

KEYWORDS: switching costs, network effects, lock-in, network externalities, co-ordination, indirect network effects

JEL CLASSIFICATIONS:
L130 market structure, firm strategy and market performance: oligopoly and other imperfect markets; monopolistic competition; contestable markets
L150 information and product quality; standardization and compatibility
L120 market structure, firm strategy and market performance: monopoly
L140 transactional relationships: contracts and reputation: networks
D430 market structure and pricing: oligopoly and other forms of market imperfection
D420 market structure and pricing: monopoly.

Acknowledgments: I am very grateful for helpful comments from David Gill, Tim Harford, Max Tse, and especially Joe Farrell, on joint work with whom these essays are based. © Paul Klemperer, 2004, 2005

* These views are personal and should not be attributed to the UK Competition Commission or to any individual Member other than myself. Furthermore, although some observers thought some of the behaviour described below warranted regulatory investigation, I do not intend to suggest that any of it violates any applicable rules or laws.
1. Network Effects

Network effects arise where current users of a good gain when additional users adopt it (classic examples are telephones and faxes). The effects create multiple equilibria and fierce competition between incompatible networks; users’ expectations are crucial in determining which network succeeds. Early choices, such as the QWERTY typewriter keyboard, lock in the market; new entry, especially against established networks with proprietary technology, is often nearly impossible. Incompatible networks can induce efficient “competition for the market”, but more often create biases and inefficiencies. Policymakers should scrutinize markets where firms deliberately choose incompatibility.

Direct network effects arise if each user's payoff from the adoption of a good, and his incentive to adopt it, increase as more others adopt it; that is, if adoption by different users is complementary. For example, telecommunications users gain directly from more widespread adoption, and telecommunications networks with more users are also more attractive to non-users contemplating adoption.

Indirect network effects arise if adoption is complementary because of its effect on a related market. For example, users of hardware may gain when other users join them, not because of any direct benefit, but because it encourages the provision of more and better software.

Extensive case studies and more formal econometric evidence document significant network effects in many areas including, for example, telecommunications, radio and TV, computer hardware and software, applications software and operating systems (including Microsoft’s), securities markets and exchanges (including Ebay), and credit cards (see, for example, Gabel (1991), Rohlfs (2001), Shy (2001), and Gandal’s (forthcoming) article in this Volume).

Usually adoption prices do not fully internalize the network effects, so there is a positive externality from adoption. A single network product therefore tends to be under-adopted at the margin – this issue was the main focus of the early literature (see, e.g., Leibenstein (1950), Rohlfs (1974)). However, if two networks compete, then adopting one network means not adopting the other which dilutes or reverses the externality.
More interestingly – and the starting point for the more recent literature - network effects create incentives to “herd” with others. In a static (simultaneous-adoption) game there are often multiple equilibria, so expectations are crucial, and self-fulfilling. Likewise a dynamic (sequential-adoption) game exhibits positive feedback or “tipping” – a network that looks like succeeding will as a result do so (see, e.g., David (1985), Arthur (1989), Arthur and Rusczcynski (1992)).

How well competition among incompatible networks works depends dramatically on how adopters form expectations and coordinate their choices. If adopters smoothly coordinate on the best deals, vendors face strong pressure to offer them. Competition may then be unusually fierce because all-or-nothing competition neutralizes horizontal differentiation – since adopters focus not on matching a product to their own tastes but on joining the expected winner.

However, coordination is not easy. With simultaneous adoption, adopters may fail to coordinate at all and “splinter” among different networks, or may coordinate on a different equilibrium than the one that is best for them - for example, each adopter may expect others to choose a low-quality product because it is produced by a firm that was successful in the past. Furthermore, consensus standard-setting (informally or through standards organizations) can be painfully slow when different adopters prefer different coordinated outcomes (see Bulow and Klemperer (1999)). Coordination through contingent contracts is possible in theory (see, e.g., Dybvig and Spatt (1983), Segal (1999)), but seems uncommon in practice.

When adoption is sequential, we see early instability and later lock-in (see, e.g., Arthur (1989)) – this corresponds to the multiple equilibria that arise with simultaneous adoption. Because early adopters influence later ones, long-term behavior is determined largely by early events, whether accidental or strategic. In theory, at least, fully sequential adoption achieves the efficient outcome if it is best for all adopters, but more generally early adopters’ preferences count for more than later adopters’: this is “excess early power”. Note that “excess early power” does not depend on “excess inertia”, that is, on incompatible transitions being too hard given ex post incompatibility. (Both “excess inertia”, and its opposite, “excess momentum”, are theoretically possible, see Farrell and Saloner (1985).)

Firms promoting incompatible networks compete to win the pivotal early adopters, and so achieve ex post dominance and monopoly rents. Strategies such as
penetration pricing and preannouncements (see, e.g., Farrell and Saloner (1986)) are common. History, and especially market share, matter because an installed base both directly means a firm offers more network benefits and boosts expectations about its future sales. Such “Schumpeterian” competition “for the market” can neutralize (or even overturn) excess early power if promoters of networks that will be more efficient later on set low penetration prices in anticipation of this (see Katz and Shapiro (1986a)). More commonly, though, late developers struggle while networks that are preferred by early pivotal customers thrive.

So early preferences and early information are likely to be excessively important in determining long-term outcomes. For example, whether or not the Dvorak typewriter keyboard is really much better than QWERTY (as David (1985) contends), there clearly was a chance in the 1800s that a keyboard superior to QWERTY would later be developed, and it is not clear what could have persuaded early generations of typists to wait, or to adopt diverse keyboards, if that was socially desirable. So it seems unlikely that the market gave a very good test of whether or not waiting was efficient. (Liebowitz and Margolis (1990) and Liebowitz (2002) contest both the details of the QWERTY example, and that network effects are significant more generally, but at least the second view is probably a minority one.)

Despite the possibility of competition for the market passing ex post rents through to earlier buyers, incompatibility often reduces efficiency and harms consumers in several ways:

Incompatibility means that consumers are faced with either a segmented market with low network benefits, or – if the market does “tip” all the way to one network – with reduced product variety and without the option value from the possibility that a currently inferior technology might later become superior. Product variety is more sustainable if niche products are compatible with the mainstream, and so don’t force users to sacrifice network effects.

These direct costs of poor coordination by adopters may be excacerbated by weaker incentives for vendors to offer good deals. For example, if a firm like Microsoft is widely believed to have the ability to offer the highest quality, it may never bother to do so: the fact that everyone expects Microsoft could recapture the market if it ever lost any one cohort of customers (or lost any one cohort of providers of complementary
products), means everyone rationally chooses Microsoft even if it never actually produces high quality or offers a low price (see Katz and Shapiro (1992)).

Ex post rents are often not fully dissipated by ex ante competition, especially if expectations fail to track relative surplus. Worse, the rent dissipation that does occur may be wasteful, such as socially inefficient marketing. At best, ex ante competition induces “bargain-then-ripoff” pricing (low to attract business, high to extract surplus) but this distorts buyers' quantity choices and gives them artificial incentives to be or appear pivotal.

Furthermore, outcomes are biased in favor of a proprietary technology (e.g., Microsoft’s) whose single owner has the incentive to market it strategically over “open” unsponsored alternatives (e.g., Linux) - see, e.g., Katz and Shapiro (1986b). As discussed above, outcomes are also often biased in favor of networks that are more efficient early on, and are generally biased in favor of established firms on whom expectations focus. The last bias implies entry with proprietary network effects is often nearly impossible (and frequently much too hard from the social viewpoint even given incompatibility). And this in turn makes it easier to recoup profits after predatory behaviour that eliminates a rival, and so encourages such predation.

So while incompatibility does not necessarily damage competition, it often does, and firms may therefore also dissipate further resources creating and defending incompatibility.

If firms offer compatible products, then consumers don’t need to buy from the same firm to enjoy full network benefits, and (differentiated) products will be better matched with customers. Consumers will be willing to pay more for these benefits, and this may encourage firms to choose compatibility. But compatibility often intensifies competition and nullifies the competitive advantage of a large installed base, whereas proprietary networks tend to make competition all-or-nothing, with the advantage going to large firms, and may completely shut out weaker firms. So large firms and those who are good at steering adopters’ expectations may prefer their products to be incompatible with rivals’ (see, e.g., Katz and Shapiro (1985), Bresnahan (2001)), and may be able to use their intellectual property to enforce this.
Competition with incompatible network effects is closely related to other forms of competition when market share is important, especially competition when consumers have switching costs (see, e.g., Klemperer (1995), Farrell and Klemperer (forthcoming), and the companion-piece to this article, Klemperer (forthcoming)), and has similar broader implications (e.g., for international trade, see Froot and Klemperer (1989)).

Because competition “for the market” differs greatly from conventional competition “in the market”, and especially because capturing consumers’ and complementors’ expectations can be so profitable, competition policy needs to be vigilant against predatory or exclusionary tactics by advantaged firms, including deliberately creating incompatibility by misusing intellectual property protection. Thus, for example, the network effect by which more popular operating systems attract more applications software took centre stage in both the US and the European Microsoft cases (see, e.g., Bresnahan (2001)). And, because coordination is often important and difficult, institutions such as standards organizations matter, and government procurement policy takes on more significance than usual.

In summary, network effects can involve efficient competition for larger units of business --- "competition for the market" --- but very often make competition, especially entry, less effective. So I, and others, recommend public policymakers should have a cautious presumption in favor of compatibility, and should look particularly carefully at markets where incompatibility is strategically chosen rather than inevitable.

Farrell and Klemperer (forthcoming) contains a recent and comprehensive survey of network effects.

Paul Klemperer,
Member, UK Competition Commission, and
Edgeworth Professor of Economics, University of Oxford.
References for Section 1


2. Switching costs

Switching costs arise when transactions, learning, or pecuniary costs are incurred by a user who changes suppliers (including for "follow-on" or "aftermarket" products such as refills and repairs). The ex-post market power that switching costs give suppliers need not create inefficiencies, and early "bargain" prices can compensate consumers for later "rip-off" pricing. More often, however, switching costs make new entry hard, distort firms’ product ranges, raise firms’ profits and lower consumer and social welfare. Similar issues arise in "shopping-cost" markets. Policymakers should scrutinise markets where firms deliberately choose incompatibility.

A product exhibits classic switching costs if a buyer will purchase it repeatedly and find it costly to switch from one seller to another. For example, there are high transaction costs in closing an account with a bank and opening another with a competitor; there may be substantial learning costs involved in switching between computer-software packages; and switching costs can also be created by non-linear pricing as, for example, when an airline enrolls passengers in a “frequent flyer” program that gives them free trips after flying a certain number of miles with that airline.

Switching costs also arise if a buyer will purchase “follow-on”, or “aftermarket”, products such as service, refills or repairs, and find it difficult to switch from the supplier of the original product. In short, switching costs are created whenever the consumer makes an investment specific to his current seller, that must be duplicated for any new seller.

Large switching costs lock in a buyer once he makes an initial purchase, so unless sellers specify all the future prices and qualities of their products, a long-term relationship is governed by short-term contracts. This creates ex post market power, for which firms compete ex ante; they use strategies such as penetration pricing, price wars, and introductory offers to fight for market share that will generate future profits (see e.g. Klemperer (1987a, 1989)).

The central question in the literature is the extent to which this fierce ex ante competition for buyers is an adequate substitute for more standard period-by-period competition without switching costs.
In the simplest models, firms’ low “bargain” prices to new customers exactly compensate buyers for the high “rip-off” prices they will pay after lock-in, so buyers’ total “life-cycle” payments are unaffected by their switching costs, and the absence of any price commitments leads to no inefficiencies. But things do not usually work so well:

Most theoretical models confirm the popular intuition that switching costs raise firms’ profits, and lower social welfare. Switching costs can segment even an otherwise undifferentiated market as firms concentrate on exploiting their established customers and do not compete aggressively for their rivals’ buyers. Unless new firms will enter the market in the future, an existing firm would usually expect to earn future profits even if it made no current sales (because it could then usually capture a few of its rivals’ lower-switching cost customers, by setting a lower price than its rivals). So rent dissipation is low, allowing oligopolists to extract positive profits overall (as illustrated by Farrell and Shapiro (1988), Padilla (1995), Chen (1997), Taylor (2003), and others). Furthermore, consumers facing switching costs care about expected future prices as well as about current prices, and are therefore generally less sensitive to current prices than absent switching costs. So firms generally have less incentive to cut prices, and prices and profits are therefore higher, than absent switching costs (see Klemperer (1987b), Beggs and Klemperer (1992)). Although it is possible to construct models in which switching costs lower profits by making special assumptions about consumers’ expectations and tastes (see von Weizsäcker (1984)), and there is little convincing empirical evidence, the limited laboratory evidence suggests switching costs raise profits (see Cason and Friedman (2002)).

Even when ex ante competition fully dissipates ex post rents, it may do so in unproductive ways such as through socially inefficient marketing. The “lo-hi” pricing patterns distort buyers’ choices – for example, if a printer is priced below cost but the locked-in ink is expensive, buyers may buy an over-specified printer but then use it too little from the social viewpoint. Such pricing also gives customers wrong signals about whether to switch. If consumers do switch, direct costs are incurred (Klemperer (1988)), and if consumers avoid those costs by not switching, that obstructs efficient matching between buyers and sellers. Product variety is less sustainable in the market than if niche
products are compatible with mainstream products and so don’t require new users to incur switching costs.

Switching costs also hamper forms of entry that must persuade customers to pay those costs, in particular, large-scale entry that seeks to attract other firms’ customers (for instance to achieve minimum viable scale, if the market is not growing quickly) (Klemperer (1987c)). The difficulty of new entry may be broadly efficient given switching costs, but is nevertheless a social cost of switching costs.

So switching costs often damage competition, and firms may therefore also dissipate further resources creating and defending incompatibilities – as, for example, when Gillette famously, and repeatedly, changed the design of its razors to prevent competing manufacturers from selling compatible blades. Likewise, it is alleged that both IBM and Microsoft have deliberately obstructed compatibility between their products and those of their competitors, and Kodak tried to prevent independent repair firms from servicing its photocopiers.

The bargain-then-ripoff pricing structure is clearest when new and locked-in customers are clearly distinguished and can be charged separate prices, for example, when prices are individually negotiated, or when locked-in customers buy separate “follow-on” products such as parts and service, rather than repeatedly buying the same good.

If, instead, each firm has to set a single price to old (locked-in) and new customers (see Beggs and Klemperer (1992)), that price must compromise between a high price to exploit current locked-in buyers, and a lower price to attract more buyers to lock in and exploit later. The implications for competition and welfare are similar to those for the bargain-then-ripoff case above, except that here switching costs also create a “fat-cat” effect: firms with large customer bases set higher prices, because they have more to gain from harvesting current customers than winning new ones.

On the one hand the “fat-cat” effect is a further force for high prices - firms price less aggressively both because they recognise that if they win fewer customers today,
their rivals will be “fatter” and therefore less aggressive tomorrow, and because consumers recognise this too and so are less impressed by lower current prices. On the other hand it actually facilitates entry that focuses purely on new customers - since an incumbent’s incentive to set a high price against its locked-in buyers creates a price umbrella under which a new entrant can come in.

The fat-cat effect means large shares tend to shrink and small shares to grow; when firms’ shares are similar, they return to stable steady state after any shock. More generally, the tradeoff between harvesting old customers and investing in new ones depends on interest rates, the state of the business cycle, expectations about exchange-rates, etc, with implications for macroeconomics and international trade (Chevalier and Scharfstein (1996), Froot and Klemperer (1989), Klemperer (1995)).

Some of the same issues as with switching costs arise when shops advertise only some of their prices: customers become “locked in” as they bear the costs of going to a shop, and only afterwards learn its other prices. Just as with dynamic switching costs, this tends to produce bargains on advertised (“loss leader”) prices and corresponding ripoffs on un-advertised prices (see Lal and Matutes (1994)).

Also closely related are “shopping-cost” markets where consumers face costs of using different suppliers for different goods in a single period but all prices are advertised (though neither time nor commitment problems are central in these markets). Shopping costs encourage firms to offer a full product line – for example, a supermarket stocks a broad range of products to encourage consumers to shop only there – and so help explain multi-product firms. Indeed firms’ product ranges may be too broad from the social viewpoint (Klemperer and Padilla (1997)), but may also be too similar to each other (Klemperer (1992)) so that there is too little variety in the market as a whole. Shopping costs also make single-product entry hard. However, the “mix-and-match” literature, beginning with Matutes and Regibeau (1988), suggests that firms typically prefer to be compatible (no shopping costs) rather than incompatible (infinite shopping costs), at least in symmetric single-period duopolies.
Other literature related to switching costs includes that on network effects (see e.g. Farrell and Klemperer (forthcoming), and the companion-piece to this article, Klemperer (forthcoming)), search costs (see e.g. Stiglitz (1989)), and “experience goods” (see e.g. Schmalensee (1982)).

The theoretical literature on switching costs described above arguably began with Selten's (1965) model of “demand inertia” (which assumed a firm's current sales depended in part on history, even though it did not explicitly model consumers' behavior in the presence of switching costs), and then took off in the 1980s with contributions from von Weizsäcker (1984), Klemperer (1983, 1987a,b,c), Farrell and Shapiro (1988), and others. But although there is an extensive empirical marketing literature on brand loyalty (or “state dependence”) which often reflects, or has equivalent effects to, switching costs (summarized in, Seetharam et al. (1999)), the empirical economics literature on switching costs is smaller and more recent than the theoretical literature.

Only a few studies attempt to directly measure switching costs. Where micro data on individual consumers' purchases are available, a discrete choice approach can be used to explore the determinants of a consumer's probability of purchasing from a particular firm (e.g., Greenstein (1993) on computer systems procurement, Shum (1999) on breakfast cereal purchases), but because switching costs are usually both consumer-specific and not directly observable, and micro data on individual consumers' purchase histories are seldom available, less direct methods of assessing the level of switching costs are often needed (e.g. Kim et al. (2003) estimate a first-order condition and demand and supply equations for Norwegian bank loans, and Shy (2002) uses data on prices and market shares for the Israeli cellular phone market).

One defect of most empirical studies is that few of them model the dynamic effects of switching costs that are the main focus of the theoretical literature; most of them assume consumers myopically maximize current utility without considering the future effects of their choices.

Nevertheless the empirical literature does suggest that switching costs play an important role in many industries including credit cards, cigarettes, supermarkets, air
travel, phone services, electricity suppliers, bookstores, and automobile insurance – see Farrell and Klemperer (forthcoming) for references, and Klemperer (1995) for more examples of markets with switching costs; as technology continues to develop, products become more complex, and services become more important, the importance of switching costs seems likely to increase further.

Because switching costs very often make competition, and especially entry, less effective, I (and many others) favor cautiously pro-compatibility public policy. Policymakers should look particularly carefully at markets where incompatibility is strategically chosen rather than inevitable. Because buyers’ early choices are often crucial, and depend on their expectations about firms’ future behavior, provision of information and consumer protection against deception, etc. are more important than usual. And competition policy must recognize that the analysis of mergers, monopolization, intellectual property, and predation, are all affected by switching costs; it is unsurprising that switching costs have featured importantly in many of the world’s best-known and most significant antitrust cases, including the IBM, the Kodak, and the European Microsoft cases.

Farrell and Klemperer (forthcoming) contains a recent and comprehensive survey of switching costs.

Paul Klemperer,  
Member, UK Competition Commission, and  
Edgeworth Professor of Economics, University of Oxford.
References for Section 2


