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

Part III of IV   Network Effects Section

December 2004

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: 2004

PRELIMINARY DRAFT: PLEASE SEND COMMENTS

1 Introduction

2 Switching Costs and Competition

3 Network Effects and Competition

3.1 Introduction

When resource scarcity drives prices, it pays to be different and avoid the crowd to improve terms of trade. But when opportunities for trade are scarce, it can pay to coordinate and follow the crowd to create more opportunities for beneficial interactions. Thus it is useful to speak English because many others do; driving is easier if everyone keeps right — or if everyone keeps left; a thicker market gives each participant more chances to trade. There are *network effects* if one agent's adoption of a good (a) benefits other adopters of the good (a "total effect") and (b) increases others' incentives to adopt it (a "marginal effect").

Classic (or peer-to-peer) network effects arise when every adoption thus complements every other, typically by giving each adopter more prospective partners. In human languages, for example, and in communications services (telephone, fax, email, instant messaging): users value the option to communicate with more others.

Such benefits are seldom uniform: a telephone user gains more when her friends adopt than when strangers do. Indeed, adoption by spammers or telemarketers *harms* other adopters and makes them *less* keen to adopt; yet a few such nuisance adopters will not overturn the overall network effect.

Indirect network effects can arise even where—in a sense—nuisance adopters are common. A new adoption will generally induce re-equilibration by other adopters. Even if A finds B a nuisance, B's adoption may encourage C to adopt, and A may want to interact with C. This can matter even for classic network effects: A knows he will never talk to B, but hopes that if B adopts it will induce some of A's friends to adopt. Unlike the classic definition, this logic assumes a (so far unspecified) degree of re-equilibration by others after B adopts.

In particular, suppose there are complementarities between two types of adopter such as buyers and sellers, so each type values increased adoption by the other type, although congestion or competition can make each adopter *dis*like more adoption by his own type. One can focus on (say) type-one adopters, with type-two adopters in the background, and evaluate the effect on each type-one adopter of incremental adoption by other type-ones, *assuming that type two re-equilibrates* given type-one adoptions.[1] Because

[1] We do not usually assume that type one then re-equilibrates, etc.; assuming full re-equilibration would unhelpfully mean that payoffs were only defined at equilibria. But note that this convention on who is taken to re-equilibrate makes the possibly arbitrary classification into "types" matter.

of the complementarity between types, more adoption by type-ones induces more adoption by type-twos, which benefits type-one adopters and makes them more willing to adopt. An **indirect network effect** arises if that "indirect" benefit outweighs any direct loss from more adoption by one's own kind. Thus an indirect network effect among type-ones satisfies the total and marginal criteria above, but does so through re-equilibration by type-twos rather than directly—perhaps even overcoming an adverse direct effect.

Often, one type of adopter is a user of a good such as computer hardware, and the other type is a vendor of a complement such as software; more users attract more vendors of the complement, making the computer (naturally bundled with the option to buy software for it) more valuable to users. This is called the hardware-software paradigm, although a leading example puts Microsoft's Windows in the role of "hardware," and applications software in the role of "software." Similarly, Visa and Mastercard are widely accepted by merchants, making them appealing to consumers, and their widespread consumer adoption encourages merchants to accept them. Such indirect network effects are common; indeed, Rochet and Tirole (2003) argue that network effects predominantly arise in this way.

Positive cross-effects or complementarities are even more common, and do not guarantee a network effect. Even in classic competitive markets, sellers like there to be more buyers, and vice versa. In perfectly competitive markets these effects are pecuniary and may cancel one another, so there is no network effect. But if larger markets are more efficient, the positive cross-effect can outweigh negative effects of own-kind adoption, along a relevant growth path. Typically we think of the growth path as involving free entry by one side (e.g. software vendors) given a growing number of users on the other side (e.g. buyers of hardware or an operating system).[2]

Network effects also arise in other ways, including pricing. If a good gets cheaper as more buyers adopt it, this creates a network effect among them, although not necessarily for society.[3] For cost-based or other reasons, a firm sometimes sets price policies that create a network effect among buyers (as when a mobile-phone provider charges less for calls to other subscribers than for calls off its network). A larger market may support additional sellers, and

---

[2]For theories of indirect network effects through improved supply in a complement see Katz and Shapiro (1985), Church and Gandal (1992, 1993), Chou and Shy (1990), and Economides and Salop (1992); Gandal (1995) and Katz and Shapiro (1994) review this literature. Liebowitz and Margolis (1994) argue that indirect network effects lack the welfare properties of direct effects; see also Clements (2003); but Church, Gandal and Krause (2003) argue otherwise.

Presumably we could have network effects with several classes of adopter, each class benefiting only from adoption by one other class, but in practice models tend to assume either classic (single-class) or indirect (two-class) cases, although multi-component systems are sometimes studied.

[3]As Liebowitz and Margolis have stressed and as we discuss below, if larger markets are not more efficient, any network effects are pecuniary (e.g., the gains to buyers from a lower price just match the losses to sellers).

thus be more competitive (Ramseyer, Rasmusen and Wiley 1991; Segal and Whinston 2000) or more productively efficient (Stigler 1951). And there may be price-mediated network effects in decreasing-cost competitive industries.

Price-mediated network effects can also reflect production-side *economies of scale* that are passed through to consumers. By no means all economies of scale are network effects, but the distinction is economic rather than technological: it depends on who gains from the scale economy. Thus, if a firm's price and quality do not respond to its scale, then buyers have no incentive to coordinate on a seller. But if prices do respond to scale then scale economies can create a network effect. For instance, if public transport is always priced at average cost, it gets cheaper the more it is used, creating a price-mediated network effect (as well as perhaps a frequency-of-service network effect) among travelers.[4] Similarly, Bagwell and Ramey (1994) and Bagwell's chapter in this *Handbook* show how economies of scale in retailing can encourage consumers to coordinate (perhaps by responding to advertising) on large retailers. And if a firm will exit without sufficient demand, and buyers face switching costs, then each consumer wants to buy a product that enough other customers will buy (Beggs 1989).

3.1.1 Outline of Section 3

Section 3.2 describes some case studies and empirical work. Section 3.3, like the early literature, asks whether a single network will be under-adopted at the margin, and whether a network effect is an externality, as the total effect can imply. But the modern literature focuses more on how the marginal effect can create multiple equilibria among adopters, making coordination challenging and giving expectations a key role. As a result, network markets often display unstable dynamics such as critical mass, tipping, and path dependence, including collective switching costs. Section 3.4 argues that coordination is central and can be hard even despite helpful institutions. Section 3.5 discusses how adoption in network markets favors the status quo; such "inertia" has important implications for competition. Sections 3.3 to 3.5 thus study adopters' collective behavior, given their payoff functions including prices. Those sections thus describe adoption dynamics when each network good is unsponsored (competitively supplied), and also describe the demand side generally, including when network goods are strategically supplied.

Section 3.6 discusses how a strategic sponsor might address coordination and externality problems; section 3.7 considers competition between sponsors of incompatible network products. In light of this analysis of incompatible competition, section 3.8 asks whether firms will "choose to compete" with compatible or incompatible products, and section 3.9 discusses public policy.

[4] If prices are fixed in the short run but set to average cost in the long run, there is a long-run network effect but no short-run network effect via price.

3.2 Empirical Evidence

3.2.1 Case Studies

Telecommunications Much early literature on network effects was inspired by telecommunications. Since telecommunications at the time was treated as a natural monopoly, the focus was mainly on how second-best pricing might take account of network effects/externalities, and on how to organize "universal service" cross-subsidies to marginal (or favored) users.[5]

Modern telecommunications policy focuses more on supporting the possibility of efficient competition. Compatibility in the form of interconnection, so that a call originated on one network can be completed on another, is fundamental to this.[6] Unlike many compatibility decisions elsewhere, it is often paid for, and is widely regulated. Brock (1981) and Gabel (1991) describe how, in early unregulated US telephone networks, the dominant Bell system refused to interconnect with nascent independent local phone companies. Some users responded by subscribing to both carriers; this "multi-homing" somewhat blunted the network effects, as do similar practices such as merchants accepting several kinds of payment cards.

Standards issues also arise in mobile telephony, although users on incompatible standards can call one another. Most countries standardized first- and second-generation air interfaces, predominantly on GSM, but the US did not set a compulsory standard for the second-generation air interface.

Radio and Television Besen and Johnson (1986) discuss standards obstacles to the adoption of AM stereo in the US after the government declined to mandate a standard; they argue that the competing standards were similar enough, and demand limited enough, for such a leadership vacuum to stall the technology. Greenstein and Rysman (2004) give a similar interpretation of the early history of 56k modem standards.

In television, governments have imposed standards, but they differ among countries; Crane (1979) interprets this as protectionist trade policy. Besen and Johnson describe how the US initially adopted a color TV standard that was not backward compatible with its black-and-white standard, so that color broadcasts could not be viewed at all on the installed base of sets; after brief experience with this, the FCC adopted a different standard that was backward-compatible. Farrell and Shapiro (1992) discuss domestic and international processes of picking high-definition television standards.

[5]See for instance Squire (1973), Rohlfs (1974), Kahn and Shew (1987), Einhorn (1993), Yannelis (2001), Barnett and Kaserman (1998), Cremer (2000), and Mason and Valletti (2001).

[6]Besen and Saloner (1989, 1994) studied standards and network effects in telecommunications; the International Telecommunications Union (ITU) has an entire "standardization sector."

Microsoft Powerful network effects arise in computer platforms including operating systems. Because they have many users, Microsoft's operating system platforms attract a lot of applications programming. An indirect network effect arises because application software writers, and other complementors, make it their first priority to work well with the dominant platform, although many applications are "ported" (a form of multi-homing), softening this effect. As we explore below, incompatible competition (and entry in particular) may well be weak *unless* applications programmers, consumers, and equipment manufacturers would rapidly coordinate and switch to any slightly better or cheaper operating system.[7]

The US antitrust case against Microsoft focused on this network effect or "applications barrier to entry", but did not claim that Windows is "the wrong" platform or that the network effect created illegal monopoly. Rather, Microsoft was convicted of illegal acts meant to preserve the network barrier against potential weakening through the Netscape browser and independent "middleware" such as Java.[8] Bresnahan (2001) argues that internal strategy documents show Microsoft executives' sophisticated analysis of network effects.

Others complain that Microsoft vertically "leverages" control from the operating system to other areas, such as applications and servers. The European Commission's 2004 order against Microsoft focused in part on leverage into media viewers and in part on interface standards between PCs and servers.

In software more generally, Shurmer (1993) uses survey data and finds network effects in word processing and spreadsheet software; Liebowitz and Margolis (2001) however argue that product quality largely explains success. Henderson and Gawer (2003) discuss Intel's response to problematic opportunities for leverage.

Computers Gabel (1991) contrasts case studies of standards in personal computers and in larger systems. In personal computers, initial fragmentation was followed by the rise of the IBM/Windows/Intel (or "Wintel") model, pushed by IBM, which meant to keep proprietary control, and tried to reassert it with its PS/2. The standard, which lets many firms complement the microprocessor and operating system (and to a lesser extent lets others, such as AMD and Linux, compete with those), has thrived, in part

[7] A barrier to incompatible entry matters most if there is also a barrier (here, intellectual property and secrecy) to compatible entry.

[8] 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. A good introduction to the case is the 2001 decision of the DC Court of Appeals. A discussion by economists involved in the case is Evans *et al.* (2000); Fisher (2000) and Schmalensee (2000) give briefer discussions; see also the *Journal of Economic Perspectives* symposium (2001). See also Evans and Schmalensee (2001), and Rubinfeld (2003). Werden (2001) discusses the applications barrier to entry. Lemley and McGowan (1998b) discuss Java. (Note: Farrell was a consultant, and later an official, for the Justice Department.)

due to the attraction of scale for applications software vendors and others, and relatedly due to the scope for specialization: see Gates, Myrhvold and Rinearson (1996), Grove (1999), and Langlois (e.g. 1990). Outside this standard only Apple has thrived.

Credit Cards From the cardholder side, a credit card system has indirect network effects if cardholders like having more merchants accept the card and don't mind having more other cardholders. The question is more subtle on the merchant side since (given the number of cardholders) each merchant loses when more other merchants accept a card. Since this negative "total effect" applies whether or not this merchant accepts the card, Katz (2001) and Rochet and Tirole (2002) show that the "marginal effect" (adoption encourages others to adopt) may hold but the total effect may fail even taking into account re-equilibration on the customer side, if card penetration is already high and total spending does not rise much with cardholding.

Network effects color inter-system competition, and dominant systems could remain dominant partly through self-fulfilling expectations, although both merchants and cardholders often "multi-home," accepting or carrying multiple cards, which weakens network effects. The biggest card payment systems, Visa and Mastercard, have been non-profit at the system level and feature intra-system competition: multiple banks "issue" cards to customers and "acquire" merchants to accept the cards. The systems' rules affect the balance between inter- and intra-system competition. Ramsey-style pricing to cardholders and merchants may require "interchange fees," typically paid by merchants' banks to cardholders' banks: see e.g. Katz (2001), Schmalensee (2002) and Rochet and Tirole (2002). But such fees (especially together with rules against surcharges on card purchases) may raise prices to non-card customers.[9]

The QWERTY Keyboard David (1985) argued that the QWERTY typewriter keyboard became dominant through "historical small events." He suggested that QWERTY remains dominant despite being inferior (at least on a clean-slate basis) to other keyboard designs, notably the "Dvorak Simplified Keyboard" (DSK). Switching costs arise because it is costly to re-learn how to type. Network effects may arise "directly" because typists like to be able to type on others' keyboards, and "indirectly" for various reasons, e.g. because typing schools tend to teach the dominant design. The subject has been controversial and confused.

Liebowitz and Margolis (1990) deny that QWERTY has been shown to be substantially inferior, claiming that the technical evidence is mixed, weak, or suggests a relatively small inferiority—perhaps a few percent, although Puffert (2003) cites a ten percent expert estimate based on recent studies.

[9]See e.g. Schwartz and Vincent (2002), Reserve Bank of Australia (2002). (Farrell testified on this topic on behalf of the Bank.)

If the penalty is small, switching (retraining) could be privately inefficient for already-trained QWERTY typists *even without* network effects. And revealed preference already indicates that few users find it worth switching given all the considerations *including* any network effects.

But new users (who would not have to re-train from QWERTY) would find it worth adopting DSK or another alternative, *if* network effects did not outweigh their clean-slate stand-alone advantages. Combined with the technical evidence, this gives a lower bound on the strength of these network effects. If most typists type for a fifth of their working time and QWERTY has a stand-alone disadvantage of 5 percent, for instance, revealed preference of new QWERTY students suggests that the network effect is worth at least one percent of earnings.[10] Yet many would doubt that network effects are terribly strong in keyboard design: most typists work mostly on their own keyboards or their employer's, and DSK training and keyboards are available (PC keyboards can be reprogrammed). We infer that even easily disparaged network effects can be powerful.[11]

But the efficiency of typing is mostly a parable; the deeper question is whether the market test is reliable. This turns out to be a subtle issue. At the top level, the question splits into two:

(a) Ex ante: did QWERTY pass a good market test when the market tipped to it? Can we infer that it was best when adopted, whether or not it remains ex post efficient now? A short-run form of this question is whether contemporary users liked QWERTY best among keyboards on offer; a long-run version is whether the market outcome appropriately took into account that not all keyboards had been tried and that taste and technology could (and later did) change.

On the short-run question, David suggests that "small" accidents of history had disproportionate effects; a prominent typing contest was won by an especially good typist who happened to use QWERTY. He suggests that the outcome was somewhat *random* and thus may well have failed even the short-run test. Liebowitz and Margolis argue that because both typing-contest and market competition among keyboards was vigorous, one can presume that the outcome served short-run contemporary tastes.

[10] Since widespread dissemination of the PC, many typists type less than this; but for most of the keyboard's history, most typing probably was done by typists or secretaries who probably typed more than this.

[11] If one were very sure that network effects are weak, one might instead infer that the clean-slate stand-alone penalty of QWERTY must be small indeed, even negative. Even aside from the ergonomic evidence, however, that view is hard to sustain. For instance, the keyboard design problem differs among languages and has changed over time, yet QWERTY and minor variations thereof have been persistently pervasive. Thus if network effects were unimportant, the evidence from new typists' choices would imply that QWERTY was remarkably optimal in a wide range of contexts. And even if QWERTY is actually the best of all designs, the many people who believe otherwise would adopt DSK if they did not perceive network effects to be bigger.

A fortiori, David presumably doubts that the market's one-time choice of QWERTY properly took long-run factors into account. Liebowitz and Margolis do not directly address the long-run question, but suggest that it shouldn't be viewed as a market failure if QWERTY won because technically superior alternatives weren't yet on the market.[12] In sections 3.5–3.7 below we discuss market forces toward contemporaneous efficiency.

(b) Ex post: As of now, would a switch be socially efficient? Many students of keyboard design believe DSK is better on a clean-slate basis. But the slate is not clean: there is a huge installed base of equipment and training. As things stand, no switch is taking place; should one? This question in turn can take two different forms.

In a *gradual switch*, new users would adopt DSK while trained QWERTY typists remained with QWERTY. This would sacrifice network benefits but not incur switching costs; it would presumably happen without intervention if switching costs were large but network effects were weak compared to DSK's stand-alone advantage. Private incentives for a gradual switch can be too weak ("excess inertia") because early switchers bear the brunt of the lost network benefits (see 3.5 below). But equally the private incentives can be too strong, because those who switch ignore lost network benefits to those who are stranded.

In a *coordinated switch*, everyone would adopt DSK at once (already-trained QWERTY typists would retrain). Thus society would incur switching costs but preserve full network effects. Because new users would unambiguously gain, already-trained QWERTY typists will be too reluctant to participate. Even if they were willing, coordination (to preserve full network benefits) could be a challenge; if they were opposed, compulsion or smooth side payments could be required for an efficient coordinated switch; of course, compulsion can easily lead to inefficient outcomes, and side payments seem unlikely to be smooth here.

Video Recordings: Betamax versus VHS; DVD and DIVX Gabel (1991) and Rohlfs (2001) argue that the VCR product overcame the chicken-and-egg problem by offering substantial stand-alone value to consumers (for "time-shifting" or recording programs off the air) even with no pre-recorded programming for rent. By contrast, RCA and CBS introduced products to play pre-recorded programming (into which they were vertically integrated), but those failed partly because they did not offer time-shifting; laser disks suffered the same fate.

Later, the VCR market tipped, generally to VHS and away from Betamax—though Gabel reports that (as of 1991) Betamax had won in Mexico. The video rental market created network effects (users value variety and convenience of programming availability, rental outlets offer more variety in a

[12] Below, we discuss what institutions might have supported a long-run market test.

popular format, and studios are most apt to release videos in such a format). The rise of these network effects hurt Sony, whose Betamax standard was more expensive (VHS was more widely licensed) and, according to some, superior at equal network size, although Liebowitz and Margolis (1994) argue not. Gabel (1991) suggests that the strength of network effects may have surprised Sony.

Cusumano, Mylonadis and Rosenbloom (1992) describe the VHS-Betamax battle. Park (2001) and Ohashi (2003) develop dynamic model of consumer choice and producer pricing for the VCR market and assess the extent to which network effects contributed to the tipping.

In the next generation of video, Dranove and Gandal (2003) and Karaca-Mandic (2003) found substantial indirect network effects in DVD adoption. Dranove and Gandal found that a preannouncement of a competing format, DIVX, delayed DVD adoption. Both papers find cross-effects such that the content sector as a whole could profitably have subsidized hardware sales, which could motivate vertical integration.

DVD players (until recently) did not record, like the laser disk product, but many households are willing to own both a VCR and a DVD player, allowing DVD's other quality advantages to drive success in a way that the laser disk could not. Again, such multi-homing blunts the network effects and can help with the chicken-and-egg problem.

Sound Recordings and Compact Disks Farrell and Shapiro (1992) argued that although prices of CDs and players fell during the period of rapid adoption, it would be hard to explain the adoption path without network effects; on the other hand, since CD players could be connected to existing amplifiers and loudspeakers, multi-homing was easy.

Gandal, Kende and Rob (2000) estimated a simultaneous-equations model of adoption in terms of price and software availability, stressing the cross-effects that would lead to indirect network effects.

Languages Human languages display classic network effects. Changes in patterns of who talks with whom presumably explain the evolution of language, both convergent (dialects merging into larger languages) and divergent (development of dialects). English is dominant, but there have been previous bandwagons such as French in diplomacy, Latin as *lingua franca* in early modern Europe, etc. Gabel (1991) argues that sacrificing efficiency for equity is not unusual in standards negotiations.

Some Americans argue for "English only" laws based on a network externality; across the border, Canadians intervene to discourage *de facto* standardization on English (Church and King 1993). As we discuss in 3.3.5 below, the net externality involved in choosing between two network goods (such as languages) is ambiguous. Of course, many people learn more than one language, but native English speakers are less apt to do so.[13]

Law Klausner (1995) and Kahan and Klausner (1996, 1997) argue that contracts and corporate form are subject to network effects (especially under common law), as it is valuable to use legal forms that have already been clarified through litigation by others, although Ribstein and Kobayashi (2000) question this empirically. Radin (2002) discusses standardization versus customization in the law generally.

Securities Markets and Exchanges Securities markets and exchanges benefit from liquidity or thickness: see Economides and Siow (1988), Domowitz and Steil (1999), Andieh (2003). When there is more trade in a particular security its price is less volatile and more informative, and investors can buy and sell without moving the market. This helps explain why only a few of the imaginable financial securities are traded, and why each tends to be traded on one exchange unless institutions allow smooth cross-exchange trading.

Not only do buyers wish for more sellers and vice versa, but this positive cross-effect outweighs the negative own-effect (sellers wish there were fewer other sellers); the difference is the value of liquidity, an efficiency gain from a large (thick) market. This fuels a network effect.

If products are differentiated, a larger network offers more efficient matches. This is the network effect behind eBay, and could be important in competition among B2B (business-to-business) exchanges (FTC 2000; Bonaccorsi and Rossi 2002). This also captures part of the value of liquidity, in that a larger market is more likely to have "coincidence of wants."

3.2.2 Econometric Approaches

Quantitative work on network effects has focused on two questions. First, it aims to estimate and quantify network effects. Second, some less formal work aims to test implications of the theory, notably the possibility of persistent inefficient outcomes.

The theory of network effects claims that widespread adoption causes high value. How can one test this? Clearly it is problematic simply to include demand for a good as an econometric predictor of demand for that good. At the level of individual adoptions, it may be hard to disentangle network effects from correlations in unobserved taste or quality variables (Manski 1993). Moreover, dynamic implications of network effects may be hard to distinguish econometrically from learning or herding.

Meanwhile, the theory predicts path dependence, which implies both large "errors" and a small number of observations (a network industry may display a lot of autocorrelation). Likewise it predicts that modest variations in

[13] Shy (1996) stresses that who learns a second language can be indeterminate and/or inefficient.

parameters will have unpredictable effects, and focuses largely on claims about efficiency, all of which makes testing a challenge. Nevertheless, some work aims to quantify these effects.

A popular hedonic approach compares demand for two products that differ in the network effects expected; the approach aims to isolate this effect from that of other quality variables. A natural proxy for expected network effects is previous sales: lagged sales or the installed base, relying on some inertia in network size. Thus Brynjolfsson and Kemerer (1996) estimated that the value of an installed base of spreadsheet users represented up to 30% of the price of the market leader in the late 1980s; similarly Gandal (1994, 1995) found a premium for Lotus-compatibility in PC spreadsheets. Hartman and Teece (1990) find network effects in minicomputers. This approach risks misinterpreting unobserved quality as network effects; but Goolsbee and Klenow (2002) find evidence of strictly local network effects in the adoption of PCs, using geographic variation to control for unobserved quality.

Another econometric approach rests on the fact that large adopters may better internalize network effects, and may care less than smaller adopters about compatibility with others. Saloner and Shepard (1995) found that banks with more branches tended to install cash machines (ATMs) sooner. Gowrisankaran and Stavins (forthcoming) also use geographic variation to estimate network effects for automated transactions by banks. Ackerberg and Gowrisankaran (2003) aim to separate consumer-level from bank-level network effects.[14]

It is easier to identify cross-effects between complementary groups, estimating how more adoption by one affects demand by the other (but recall that complementarities need not imply network effects). Rosse (1967) documented that newspaper advertisers pay more to advertise in papers with more readers, although news readers may not value having more advertisements; by contrast, readers do value having more advertisements in the Yellow Pages (Rysman 2003). Dranove and Gandal (2003) and Karaca (2003) also focus on the cross-effects.

Testing the central efficiency implications of the theory is hard, because (a) it is hard (and not standard economic methodology) to directly assess the efficiency of outcomes, and (b) the theory's prediction that outcomes depend sensitively on early events and are insensitive to later events, costs and tastes, is also hard to test. Liebowitz and Margolis (2001) argue that software products succeed when measured quality is higher, and that prices do not systematically rise after the market tips; they infer that network effects seem unimportant.[15] Bresnahan (2004) argues that effective competition

[14] See also Guiborg (2001) and Kauffman and Wang (1999).

[15] They (1994) also argue that network effects may be essentially exhausted at relevant scales, so that the u function flattens out, as Asvanund et al. (2003) found in file sharing. However, Shapiro (1999) argues that network effects are less likely than classic scale economies to be exhausted.

for the market occurs only at rather rare "epochs" or windows of opportunity, so that high quality may be necessary but is not sufficient for success.

Fascinating though they are, these case studies and empirics do not satisfyingly resolve the theoretical questions raised below, which concern the efficiency of equilibria.

3.3 Under-Adoption and Network Externalities

In this sub-section we follow the early literature on network effects in focusing on the single-network case and on the total effect or (often) adoption externality.

3.3.1 Formalities

Each of K players, or adopters, chooses an action: to adopt a product or not, or to adopt one product (network) or another. We often interpret these players not as individuals but as "groups" of adopters, where group i is of size n_i and $\sum n_i = N$. Often (but see 3.3.2), we treat each group as making an all-or-nothing choice.

Player i has payoff $u_a^i(x)$ from action a if a total of x adopters choose action a; for simplicity, assume there is only one alternative, a'.[16] Recalling our definition in 3.1, **we say that there are** *network effects* **in** a **if, for each** i**, both the payoff** $u_a^i(x)$ **and the adoption incentive** $u_a^i(x) - u_{a'}^i(N-x)$ **are increasing in** x.

For simplicity, the literature often takes $K = 2$, though the problems might not be very interesting with literally only two adopters. Figure 1 illustrates two groups' choices to adopt a single product or not. If non-adopters' payoffs are unaffected by how many others adopt, then we can normalize the payoff from non-adoption as zero, and (dropping the subscript) write $u^i(x)$ for i's payoff from adoption, as in **Figure 1**:

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

Network effects arise for this single product if $u^i(N) > u^i(n_i)$ for $i = 1, 2, ...K$;[17] in 3.3.3, we show that this implies both parts of our definition. However, often the leading alternative to one network product is another, as in **Figure 2**:

[16] It is not immediately clear how best to extend the definition to more than two alternatives: for which alternative(s) a' must the "adoption incentive" described in the text increase with adoption of a, and which alternatives does that adoption displace? The literature has not focused on these questions and we do not address them here.

[17] We often assume for clarity that u is *strictly* increasing when there are network effects.

	Group 2 adopts A	Group 2 adopts B
Group 1 adopts A	$u^1_A(N), u^2_A(N)$	$u^1_A(n_1), u^2_B(n_2)$
Group 1 adopts B	$u^1_B(n_1), u^2_A(n_2)$	$u^1_B(N), u^2_B(N)$

Network effects arise then if $u^i_a(N) > u^i_a(n_i)$ for $i = 1, 2$ and $a = A, B$; again, this implies both parts of our definition.

Network effects are *strong* if they outweigh each adopter's preferences for A versus B, so that each prefers to do whatever others do. Then "all adopt A" and "all adopt B" are both Nash equilibria of the simultaneous-move noncooperative game whose payoff matrix is Figure 2. Strong network effects thus create multiple equilibria if adoption is game-theoretically simultaneous (i.e., players cannot react to one another's actual choices but must base their decisions on expectations). For a single network product (Figure 1), network effects are strong if, for all i, $u^i(N) > 0$ (each would adopt if others do) and $u^i(n_i) < 0$ (each will not if others do not). Thus "no adoption" can be an equilibrium even for valuable network goods: the *chicken-and-egg problem*.

3.3.2 What are the Groups?

Like much of the literature, we assumed above that each group of adopters is a single player in a game, and that each adopter within the group values the total number of adopters.[18] But with indirect network effects, relaxing the assumption that groups are single players can help analyze complementary adoption by different kinds of adopter. Then we might call each kind of adopter a group: for instance, in camera formats, we might make photographers one group and film processors the other. Then each group's benefit from adoption increases when the other group adopts more strongly. Usually (as here) this greatly reduces the number of groups.

It also departs from our formal definition in two ways. First, each group does not coordinate internally and does not make an all-or-nothing adoption choice; rather, some but not all members of each group adopt. Second, there may be no intra-group network effects, and may even be intra-group congestion effects. Thus, holding fixed the number of photographers, a developer prefers fewer other developers for competitive reasons; compare also our discussion of merchants accepting credit cards.

One approach views only photographers as adopters, and diagnoses an "indirect network effect" among them, mediated through the equilibrium response of film processors. Doing so returns us to the strict framework above, but pushes the processors into the background, losing detail and insight about each side's response to the other.

Note also that adoption choices can be made at several different vertical

[18] In reality network benefits are not homogeneous (Beige 2001 discusses local network effects, or communities of interest). Also note that if $u^i_a(x)$ is linear in x and independent of i, then the total value of the network is quadratic in x: "Metcalfe's law." Swann (2002) and Rohlfs (2001) argue that this is very special and even extreme.

levels (see section 3.8.3b). For instance, in the PC industry, DRAM memory technology is arguably chosen to some extent by memory manufacturers, by producers of complements such as chipsets, by computer manufacturers (OEMs), and/or by end users or their employers. We often think of the key "adopters" as end users choosing between standards if vendors have chosen incompatible technologies; equally, we might think of them as the vendors. Confusion can arise if end users want to be compatible but vendors do not; in the latter case, competitive issues arise (see 3.8).

3.3.3 Total and Marginal Effects

Our definition of network effects requires that (a) one agent's adoption of a good benefits other adopters (a *total effect*), and that (b) his adoption increases others' incentive to adopt (a *marginal effect*). These two conditions are logically independent. We noted above that the marginal effect might apply to merchants' decisions to accept credit cards even if the total effect does not, if a merchant's adoption hurts his rivals who don't adopt more than it hurts those who do. On the other hand, the total effect can apply where the marginal effect does not: if one firm in a standard Cournot oligopoly chooses a lower output, it benefits other firms who have chosen a low output, but those other firms then typically have an incentive to *increase* their output.[19]

Although the two conditions are logically separate, definitions in the literature often mention only the total effect. The (seldom explicit) reason is that if the total effect holds *for both alternatives* A and B then the marginal effect follows. Group 2's incentive to adopt A rather than B is $u_A^2(N) - u_B^2(n_2)$ if group 1 has adopted A; it is $u_A^2(n_2) - u_B^2(N)$ if group 1 has adopted B. The marginal effect therefore holds if $u_A^2(N) - u_B^2(n_2) > u_A^2(n_2) - u_B^2(N)$, or $u_A^2(N) + u_B^2(N) > u_A^2(n_2) + u_B^2(n_2)$; but this follows from adding the two total-effect conditions $u_y^i(N) > u_y^i(n_i)$ for $i = 2$ and $y = A, B$. In particular, if (as in Figure 1) each group's payoff from B is independent of others' choices, then there are network effects in A if and only if the total effect holds for A.

As we will see below, the total effect was central in the early literature, but more recent work depends more on the marginal effect, which is essentially Segal's (1999) "increasing externalities" or Topkis' (1978, 1998) "supermodularity" (see also Milgrom and Roberts 1990). The early literature focused on a single network, with a scale-independent outside good, and thus, although writers generally stressed the total effect, the marginal effect followed. Modern literature focuses on competing networks, which as we discuss below strengthens the marginal effect but weakens and can reverse the total effect.

[19]Other firms would have an incentive to reduce their outputs in the less-usual case that firms'outputs are "strategic complements" (in the terminology introduced by Bulow, Geanakoplos, and Klemperer (1985)). In this case, the marginal effect does also apply.

3.3.4 Under-Adoption of a Single Network

A network good is apt to be under-adopted in two ways. First, there is the "chicken-and-egg" problem. Second, if the network effect is an externality (see below), there is too little incentive to adopt at the margin. Each adopter assesses his own benefit from adoption and may not take into account the benefit to other adopters. We discuss the marginal externality here and the chicken-and-egg problem in 3.4.2 below.

In Figure 1, if $u^1(N) > 0 > u^2(N)$ then player 1 would like the "all adopt" outcome but, even if he adopts, player 2 will not. If $u^1(n_1) < 0$ then the unique equilibrium is no adoption; if instead $u^1(n_1) > 0$ then equilibrium is adoption by group 1 alone. In either case, if $u^1(N)+u^2(N) > \max[u^1(n_1), 0]$ then adoption by all would increase total surplus. Player 2 has the wrong "marginal" adoption incentives (as may player 1 also).

Sometimes wrong incentives cause no wrong choices: the efficient outcome may be an equilibrium. Say that preferences are *similar* if the players agree on the best outcome, so $u^i(N)$ has the same sign for all i. Then the efficient outcome is either "all adopt" or "no adoption," and it is an equilibrium of the simple simultaneous-adoption game suggested by Figure 1, as Liebowitz and Margolis (1994) noted. Moreover, while this equilibrium need not be unique, it is each player's best feasible outcome, and many institutions (including side payments, sequential moves and commitment, and communication) preserve and reinforce it.

But generically (when preferences are not similar) the wrong incentives will cause wrong choices. In a static framework, the externality generically makes the network too small.[20] If adoption is dynamic, for instance if costs fall over time, the same logic makes adoption too *slow*.[21] Consider a marginal adopter for whom the cost of service exceeds his private willingness to pay, but is less than that value plus the increase in others' adoption value. It is efficient to subsidize such an adopter, but it is unlikely to happen with simple linear prices. But the basic point does not depend on uniform (or even imperfectly discriminatory) prices. With complete information and adopter-specific pricing, Segal (1999) finds that because there are no externalities on non-traders, efficiency results if the sponsor simultaneously and publicly makes an offer to each adopter, but because there are positive externalities on efficient traders, there is too little adoption when offers are "private" (the sponsor does not commit to how he will trade with others).

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

[21] Dynamic adoption paths with falling prices or other "drivers" of increasing adoption have been studied by (e.g.) Rohlfs (1974), Farrell and Shapiro (1992, 1993), Economides and Himmelberg (1995), Choi and Thum (1998), and Vettas (2000). Prices may fall over time because of Coasian dynamics: see 3.6. Adoption paths can also be driven by the strengthening of network effects: human languages with more trade and travel; computer programming languages with more modularity and re-use; VCRs with more movie rental.

3.3.5 Are Network effects Externalities?

Network effects often involve externalities, in the sense that prices don't fully incorporate the benefits of one person's adoption for others. Indeed, early literature often simply called network effects "network externalities." But network effects are not always externalities, as Liebowitz and Margolis (1994) have stressed.

First, network effects can be pecuniary, if adoption just affects price but not efficiency. Then, if Figure 1 describes payoffs to buyers, sellers bear an equal negative effect. Although *buyers jointly* could be made better off by a well-targeted small subsidy from inframarginal to marginal buyers, no such subsidy can make *everyone* (sellers included) better off. Liebowitz and Margolis (1994) argue that many indirect network effects are pecuniary and do not affect economic efficiency at the margin. But Church, Gandal and Krause (2002) find a real efficiency gain when a larger "hardware" network attracts more variety of "software," not just a lower price. Moreover, if pecuniary network effects stem from decreasing costs in a competitive industry, that often reflects a real economy of scale (perhaps upstream), so there is an efficiency (not just pecuniary) benefit of coordination.

Second, *any* economic effect is an externality only if not internalized. A network effect might be internalized through side payments among adopters, although this will be hard if there are many players or private information. Alternatively (see 3.6-3.7) a seller who can capture the benefits of a larger network might internalize network effects and voluntarily subsidize marginal adopters, as in Segal's (1999) model of public offers. But unless a seller can accurately target those adopters, subsidy is costly, and while it may sometimes work well enough, it seems clear that it often won't. Indeed, first-best pricing would require the price to each adopter to be equal to incremental cost less his external contribution to others, and such pricing jeopardizes profits and budget balance. Suppose for instance that a good will be supplied if and only if all K groups agree. For first-best adoption incentives, the price facing group i should be equal to the cost C of supplying the good to all, less the additional surplus accruing to groups other than i as a result of group i's agreeing: $p_i = C - \sum_{j \neq i} u^j(N)$. Hence $\sum p_i - C = (N-1)[C - \sum u^i(N)]$, so costs are covered if and only if adoption is inefficient! (First-best incentives require that each adopter be a residual claimant, which gives the vendor a negative equity interest.) Adoption prices will often not fully internalize network effects, and a profitably supplied single network good will be under-adopted.

Third, any externalities are smaller and ambiguous when networks compete. To illustrate, suppose that $K = 3$, and that groups 1 and 2 have adopted A and B respectively; now group 3 is choosing. A-adopters (group 1) gain if group 3 adopts A, but B-adopters gain if it adopts B. When each choice means rejecting the other, the net effect on others is ambiguous.[22]

[22]To quantify, treat K as large, and approximate the set of adopters with a continuum.

Thus, as attention shifts from pricing of a monopoly network, to competition between networks, it becomes less clear where there are externalities in the standard marginal sense. But other concerns arise, as we will see.

3.4 The Coordination Problem

When networks compete, we just noted that any conventional externality becomes weaker and ambiguous. The same logic, however, *strengthens* the marginal effect—the fact that adoption encourages others to adopt the same network. A user's adoption of A instead of B not only directly makes A more attractive to others but also makes the alternative, B, less so.[23] For instance, part of the positive feedback in the adoption of CDs was the declining availability of LP records as CDs became more popular.

Through the marginal effect, strong network effects create multiple adoption equilibria and hence coordination problems. Optimal coordination is hard, as everyday experience and laboratory experiments (Ochs 1995; Gneezy and Rottenstreich 2004) confirm. Coordination problems include actual breakdowns of coordination (3.4.1) and coordination on the wrong focal point (3.4.2). Coordination is especially difficult—and the institutions to aid it work less well—when (as in the Battle of the Sexes) the drive for coordination is mixed with conflict over what to coordinate on.

3.4.1 Coordination Breakdowns: Mistakes, Splintering, and Wait-and-See

Coordination "breaks down" when adopters choose incompatible options but would all prefer to coordinate. This can happen in at least two ways, which we call confusion and splintering. Economic theorists' equilibrium perspective pushes them toward (probably over-) optimistic views on the risks of such failures, but case studies and policy discussion often implicate coordination failures.

A small shift of $\mathrm{d}x$ users from a network of size x_A to one of size x_B has a net effect on other adopters of $e = [x_B u'_B(x_B) - x_A u'_A(x_A)]\mathrm{d}x$: this has ambiguous sign and is smaller in magnitude than at least one of the $x_i u'_i(x_i)\mathrm{d}x$.

The incentive to "splinter" from what most others are doing is too strong at the margin (defection imposes a negative net externality, or conformity confers a positive externality) if $e < 0$ whenever $x_B < x_A$. When the goods are homogeneous except for network size, that condition is that $xu'(x)$ is increasing: see Farrell and Saloner (1992). In the convenient (if unrealistic) Metcalfe's Law case $u(n) = vn$, there is thus too much incentive to defect from a network to which most players will adhere. Then there is not just a benefit but a positive externality from conformity.

[23] With a continuum of adopters, the gain in A's relative attractiveness from a small increase in its adoption at B's expense is proportional not just to $u'_A(x_A)$, as it would be if A were the only network good, but to $u'_A(x_A) + u'_B(x_B)$. Note that this strengthening of the marginal effect depends on the total effect in both A and B.

Confusion First, coordination can break down by mistake or sheer confusion, because adopters do not know what others are doing.[24] Common knowledge of plans averts such confusion, and the simplest economic models assume it away by focusing on pure-strategy equilibrium, in which by definition players know one another's strategies and do not make mistakes.[25] Other models use mixed-strategy equilibrium,[26] which may be too pessimistic about coordination: each player's attempt to coordinate with others is maximally difficult in mixed-strategy equilibrium.[27]

Splintering Second, coordination can break down even in pure-strategy equilibrium with strategic certainty. This happens if product differentiation discourages unilateral moves (e.g. to slightly larger networks) but is weak enough that a coordinated move of everyone on networks B, C and D to network A would benefit all.

When there are just two networks A and B splintering is impossible if the users of each network can optimally coordinate as a group, but can arise if, for example, a coordinated move of everyone on network B to network A would benefit all of them, but the users of B cannot coordinate.

The incompatible outcome is thus (in game-theory language) an equilibrium but not coalition-proof: if multiple decision makers could coordinate a move they would all do better. We call this *splintering*: a dysfunctional equilibrium with multiple small and consequently unsuccessful networks instead of one large and successful one. Common knowledge of plans does *not* avert these problems; their solution requires a leadership-like ability to focus on "let's all do X instead."

Evidence that splintering is important includes the demand for consensus compatibility standards, which provide just such leadership.[28] Such standards (see 3.4.3) go beyond mere communication of plans, since common knowledge need not cure the problem. For instance, following Thompson (1954), Hemenway (1975) and Gabel (1991) argue that early twentieth-century standardization of auto parts mainly reduced spurious variety. Even before the standardization meetings any manufacturer could have chosen to match another's (non-proprietary, non-secret) specifications; apparently

[24]In *The Gift of the Magi*, a famous short story by O. Henry, Jim sold his watch to buy his wife Della a comb; Della sold her hair to buy Jim a watch-chain. Their plans were secret because each was meant as a Christmas surprise for the other.

[25]Rationalizability, on the other hand, unhelpfully permits any outcome in a simultaneous-adoption game with strong network effects.

[26]See for instance Dixit and Shapiro (1986), Farrell (1987), Farrell and Saloner (1988), Bolton and Farrell (1990), Crawford (1995).

[27]But mixed-strategy equilibrium can be defended as a shorthand for a symmetric Bayesian-Nash equilibrium with incomplete information.

[28]An optimistic view would be that consensus standards promptly solve the problem wherever it arises, so splintering never persists. But finding consensus standards seems slow and painful, which casts doubt on such optimism. If the pain and slowness arises from difficulty in finding Pareto-improving coordinated shifts, however, then the theory sketched in the text is incomplete.

such a unilateral move would not pay, but a coordinated voluntary move did.[29] But consensus standards generally are non-binding and do not involve side payments, so they would not affect a failure to standardize that was a coalition-proof equilibrium reflection of (say) differences in tastes.

There is little theoretical work on splintering, although Kretschmer (2004) explores how it can retard innovation when there are multiple alternatives to a single established standard.[30] But it features prominently in case studies. Postrel (1990) argued that quadraphonic sound in the 1970s failed because competing firms sponsored incompatible quad systems and because hardware producers did not adequately manage complements (recorded music). Rohlfs (2001) describes how competing incompatible fax systems (invented in 1843) stalled for 140 years until consensus standardization in the late 1970s.[31] Augereau, Greenstein and Rysman (2004) claim that the adoption of 56K modem technology in aggregate was stalled by the coexistence of two equally good incompatible standards until the ITU issued a third that became focal. Saloner (1990) discusses splintering among Unix implementations (widely blamed for slow adoption of Unix until Linux became relatively focal). Besen and Johnson (1986) argued that AM stereo was adopted slowly because there were competing, broadly comparable, standards and no player could start a strong bandwagon: adopters (radio stations) avoided explicit coordination because of antitrust fears, and the FCC did not take a lead. Microsoft was accused of "polluting" or intentionally splintering the Java standard when it perceived the latter as a threat to its own non-Java standard. Rysman (2003) notes that competition in yellow pages may involve splintering, thus reducing network benefits (although he finds that this does not outweigh losses from monopoly). He does not assess whether advertisers and users might instead all coordinate on the directory that offers them jointly the best deal—a sunnier non-splintering view of incompatible competition that theory has tended to find focal.

Do similar splintering concerns arise with traditional economies of scale? In terms of cooperative game theory (how much surplus is generated by various groups of participants) network effects and economies of scale are isomorphic, so concerns about splintering parallel classic concerns about inefficiently small-scale production in monopolistic competition. Modern models of the latter, since Spence (1977) and Dixit and Stiglitz (1977), mostly attribute splintering among monopolistically competitive firms to horizontal product differentiation, and because variety is valuable, these models find that although each firm is too small to minimize average cost, it need not be too small for overall efficiency. But the classical suspicion that equilibrium involves too much fragmentation re-surfaces in that a popular claimed

[29]The point is not that there are increasing returns in compatibility benefits, but that a critical mass may be necessary to overcome differences in tastes, beliefs, etc.

[30]Goerke and Holler (1995) and Woeckener (1999) also stress inefficiencies of splintering.

[31]Economides and Himmelberg (1995) estimated a demand system for the adoption of fax under a single standard.

efficiency motive for horizontal mergers is achieving more efficient scale.[32]

Fear of Breakdowns Even mere fear of coordination breakdowns may delay adoption as people wait to see what others will do.[33] This can inefficiently slow adoption through strategic uncertainty rather than because of the externality from adoption.

3.4.2 Coordinating on the Wrong Equilibrium

Because coordination is hard, clumsy cues such as tradition and authority are often used. Schelling (1960) suggested that two people wishing to meet in New York might well go, by tradition, to Grand Central Station at noon. Many species of animals meet at fixed times or places for mating. Human meetings, and work hours, are often arranged in advance, and thus do not respond sensitively to later-revealed information about what is convenient for participants. The persistence of such clumsy solutions testifies to the difficulty of flexible optimal coordination. It would therefore be surprising if multiple adopters of networks always coordinated on the choice that gives them the most surplus.

Other parts of economics have studied the possibility of (perhaps persistent) coordination on the wrong equilibrium. Rosenstein-Rodan (1943) and Murphy, Shleifer and Vishny (1989) suggested that industrialization requires a "Big Push" that coordinates many sectors of the economy and that may not happen under *laissez-faire*. Modern economic geography sees patterns of development as partly fortuitous (Saxenian 1994, Krugman 1991, Davis and Weinstein 2002). Macroeconomists have studied how otherwise irrelevant "sunspot" signals can guide economies to good or bad equilibria;[34] game theory has studied how cheap talk can do so.[35]

Starting from a bad equilibrium, there would (by definition) be joint rewards for getting to a better equilibrium, but no rewards to individually deviating. As Liebowitz and Margolis (1994, 1995) stressed, this can suggest a role for an entrepreneur: in 3.6–3.8 below, we note some entrepreneurial tactics.

Single Network In Figure 1, if network effects are strong, so that $u^i(N) > 0 > u^i(n_i)$, then both "all adopt" and "no adoption" are equilibria, Leiben-

[32]Many people are skeptical of this (e.g., Farrell and Shapiro (2001)). A merger removes all competition between firms, whereas a common standard replaces incompatible competition with compatible competition; see 3.9.

[33]Kornish (2003) describes a "decision-theoretic" model of adoption timing under strategic uncertainty, but takes as given the behavior of all agents but one.

[34]See e.g. Cooper (1999), Cooper and John (1988), Cooper and Haltiwanger (1988), Diamond (1982), and Bryant (1994).

[35]See e.g. Farrell and Rabin (1996).

stein's (1950) chicken-and-egg problem.[36] There may also be intermediate (pure-strategy) equilibria.

With a single network, voluntary adoption is weakly Pareto-improving, so an equilibrium with more (in a set-inclusion sense) adoption Pareto-dominates one with less. Dybvig and Spatt (1983) show that there is a *maximal equilibrium*, in which all players who adopt in any equilibrium adopt. This maximal equilibrium is Pareto preferred to all other equilibria, which thus have too little adoption.[37]

The key is *expectations*. If players expect others to hang back and not adopt, they too will not adopt. Dynamic implications include positive feedback and "critical mass:" once enough adoption happens or is confidently foreseen, further self-reinforcing adoption follows. Similarly lack of adoption is self-reinforcing: a network product can enter a "death spiral" if low adoption persuades others not to adopt.[38]

This chicken-and-egg problem is a quite different form of under-adoption from the marginal externality in 3.3.4–5 above. The marginal problem arises only when preferences between equilibria are not similar,[39] could typically be helped by small subsidies to marginal adopters, and cannot be solved by voluntary joint action without side payments; whereas the chicken-and-egg problem arises even with identical adopters, might be solvable by coordinating better without side payments, and often cannot be helped by small subsidies.

Competing Networks Similar coordination problems can cause the adoption of the wrong network good. In Figure 2, if players expect others to adopt A, they will do so, but expectations in favor of B are equally self-fulfilling. And if expectations clash, so too will choices. What, then, drives expectations? In general one must look to cues outside Figures 1 and 2: below we discuss some relevant institutions.

Clumsy coordination can also blunt competitive pressures among networks, since business does not reliably go to the best offer. We discuss this further in 3.7 below.

---

[36] We use this metaphor for all such self-sustaining multiple equilibria; some might argue that it is more apt for indirect than for direct network effects (chickens, roosters, hens and eggs might all exist, or not, but are separate groups).

[37] This is characteristic of games with "strategic complements" (in the terminology introduced by Bulow, Geanakoplos, and Klemperer (1985)) or supermodularity (see Topkis (1978, 1998) or Milgrom and Roberts (1990)).

[38] This is a dynamic version of the chicken-and-egg problem. Schelling (1978) describes such dynamics in a wide range of applications. Of course, the dynamics can also work in the other direction, with critical mass and take-off. Jeitschko and Taylor (2001) study the stability of "faith-based coordination."

[39] This assumes, as does most of the literature, that each adopter's choice is zero-one. When each adopter makes a continuous quantity choice a marginal problem arises even if preferences are identical.

3.4.3 Cheap Talk and Consensus Standards

A natural response to a coordination problem is to talk. Credible talk can make plans common knowledge and thus avert confusion-based coordination failures, and may help adopters coordinate changes in plans and thus escape splintered equilibria or coordination on the wrong focal point. In fact, many voluntary "consensus standards" are reached through talk, sometimes mediated through standards organizations; David and Shurmer (1996) report that consensus standardization has grown dramatically.[40] Large official organizations often have formal procedures; smaller consortia may be more flexible.[41] The economics literature on consensus standards seems less developed than that on de facto or bandwagon standards, perhaps because reaching consensus seems political rather than a narrowly economic process.

The central observation of the game theory literature on cheap talk is that talk works less well the more conflict there is. Conflict is common in standards discussion when players want to coordinate but disagree over what to coordinate on, as in the Battle of the Sexes. For example, conflict is likely between those who are more locked in to an old technology and those who are less so, when a promising new technology arrives. Gabel (1991) argues that conflict is likely between those who are and are not vertically integrated. Conflict may also arise because active participants in standards organizations tend to be competing vendors rather than end users. Indeed, vested interest may motivate participants to bear the costs of participating.[42] Vested interest may be especially strong when potential standards incorporate intellectual property.

As a result, attempts to coordinate through talk may induce bargaining delays that dissipate much of the gains from coordination. The economics literature's stress on this observation echoes concerns of many standards participants. Economists have modeled the process as a war of attrition: participants who favor standard A hope that those who favor B will give up rather than delay further. Farrell and Saloner (1988) introduced such a model with complete information and two participants, and compared "committee" versus "bandwagon" standardization, and against a hybrid mechanism.[43] Far-

[40] Some practitioners reserve the term "standard" for formal consensus coordination. Standards organizations 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.

[41] On the institutions, see e.g. Hemenway (1975), Kahin and Abbate (1995). On the economics of consensus standards development see also Besen and Saloner (1988). Cargill (1989) describes the standards process in information technology.

Weiss and Sirbu (1990) econometrically study technology choices in voluntary consensus standards committees. Lehr (1995) describes consensus standardization in the internet; see also Simcoe (2004).

[42] Weiss and Sirbu (1990); Farrell (1996).

[43] In a hybrid mechanism, compatibility may result either by consensus or by one proponent driving a market bandwagon (but if both try simultaneously, the result is incompat-

rell (1993) and David and Monroe (1994) observe that when there is private information about the quality of proposed standards, the war of attrition may select for good proposals, although at a high cost (Simcoe 2004 shows how similar results can emerge from rational search by interested parties). They then assess efficiency consequences of rules in the consensus standards process. For instance, many standards organizations limit the exploitation of intellectual property embodied in standards (Lemley 2002), and this may reduce delays as well as limit patent-holders' *ex post* market power. Simcoe (2004) analyzes data from the Internet Engineering Task Force and finds evidence that more vested interest (measured as more patents, or more commercial participation) causes more delay. Another response is to seek rapid consensus before vested interest develops; but this can lead to excessive haste.

With two players (as in those models), either can ensure immediate consensus by conceding. With more players, Bulow and Klemperer (1999) show that delays can be very long if conceding brings no reward until others also concede, as is the case if (as in many standards organizations) a standard requires near-unanimous consensus.[44]

3.4.4 Coordination through Sequential Choice

Game theory claims that with full information and strong network effects, fully sequential adoption ensures coordination on a Pareto-undominated standard. The argument (Farrell and Saloner 1985) is fairly convincing with two groups. For simplicity, 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 of the simultaneous-adoption game. If group 1 first adopts, then group 2 will also adopt: knowing this, group 1 can (and therefore will) get $u^1(N)$ by adopting. By moving first, group 1 can start an irresistible bandwagon: it need not fear that adoption will give it only $u^1(n_1)$; thus only the efficient equilibrium is subgame-perfect when adoption is sequential.

The argument extends in theory to any finite number of players, and to the choice between two (or more) networks.[45] But it is much less compelling with many players: it assumes that each adopter sees all previous choices before making his own, and assumes strong common knowledge of preferences

ibility). Thus the consensus standards process competes directly against the bandwagon process; Gabel (1991) stresses that network effects can be realized through consensus, bandwagons, or other means. Besen and Farrell (1991) note a different form of competition among processes: less-formal consensus processes may act faster than more formal ones such as the International Telecommunications Union (ITU); Lerner and Tirole (2004) study forum-shopping equilibria in consensus standards.

[44] By contrast, they show that if a player can cease to bear delay costs by unilaterally conceding (as in oligopolists competing to win a natural monopoly), a multi-player war will quickly collapse to a two-player one. Political scientists analogously have Duverger's Law, a claim that most elections will have two serious candidates.

[45] Farrell and Saloner (1985) also show (with two groups) that cheap talk need not help when information on preferences is incomplete; Lee (2003) extends this to K groups.

and of rationality to forge a chain of backward induction with (on the order of) K steps, an unreliable form of reasoning (empirically) when K is large. Thus the theoretical result is surely too strong: the first player shouldn't count on it if $u^1(n)$ is very negative for small n; and if players won't rely on the result, it becomes false. But it does express one possible route out of inefficient coordination traps: an influential adopter could try to start a bandwagon. In this respect influence is related to size: when a big player moves, it shifts others' incentives by more than when a small player moves. Indeed, it may even become a dominant strategy for others to follow, surely a stronger bandwagon than relying on backward induction in the subgame among the remaining players. Thus size confers leadership ability, and highly concentrated markets are less apt (other things, notably conflict, equal) to be caught in pure coordination traps.

Holmes (1999) discusses the role of large players in the geographic shift of the US textile industry; Bresnahan (2001) discusses AOL's role (as a large and potentially pivotal user) in the Netscape-Microsoft battle for browser share. It is worth noting that leadership out of a coordination trap might happen at any layer of a value chain.

This result is optimistic about the ability of adoption bandwagons to avert Pareto-inferior outcomes. As we see next, however, bandwagons may be less good at balancing early and late adopters' preferences.

3.5 Inertia in Adoption

Individual switching costs can cause problems, as in section 2 above, but at least each user makes his own choice. Network effects, by binding together different users' choices, might generate a stronger and more worrying form of inertia, locking society in to an inefficient product (or behavior) because it is hard to coordinate a switch to something better but incompatible—especially where network effects coexist with individual switching costs. In a range of cases, including QWERTY, English spelling, VHS, and many computer software products, some suggest that a poor standard *inefficiently* persists because network effects create excessive inertia. Liebowitz and Margolis (1990, 1995) are skeptical (notably in QWERTY) and argue (2001) that success in computer software has followed trade reviewers' assessments of product quality; but Bresnahan (2003) argues that this has been true only in wide-open periods and that high quality is necessary but not sufficient for success. It is hard to test *ex post* excess inertia in case studies by directly assessing the efficiency of outcomes; we focus instead on the economic logic. Here we ask how much inertia there is in adoption dynamics at given prices. In 3.6–3.7, we ask how sponsors' price and other strategies affect it.

3.5.1 Ex Post Inertia

Inertia arises *ex post* if later adopters remain compatible with the installed base even though an alternative would be better if network effects were neutralized. Just as contestability theory observes that economies of scale alone do not create an advantage to incumbency, so too network effects alone need not generate inertia: in principle everyone could instantly shift to coordinate on the better alternative. But there are usually some sunk costs or switching costs; and if expectations centre on the status quo then inertia results even if there are no tangible intertemporal links.

Inertia surely is often substantial: Rohlfs (2001) argues from the history of fax that a network product without stand-alone value must be "truly wonderful and low-priced" to succeed; he and others attribute the VCR's success to its offering stand-alone value; Shapiro and Varian (1998) quote Intel CEO Andy Grove's rule of thumb that an incompatible improvement must be "ten times better."

Inertia can be efficient: incompatibility with the installed base is a real social cost if the status quo has network effects. But inertia is *ex post* "excess" if it would be more efficient for later adopters to switch, *given* earlier adopters' choice. (As that phrasing suggests, we follow the literature in assuming here that the installed base will not switch; if it would, then later adopters would sacrifice no network benefits and would collectively have excessive incentives to switch.) For example, it would be *ex post* excess inertia if society should switch to the DSK typewriter keyboard, counting the full social costs, but network effects and switching costs *inefficiently* prevent this. This requires that pivotal movers inefficiently fail to move, because they expect others not to move (the "horses" problem), or because they bear a larger share of the costs than of the benefits of moving (the "penguins" problem).[46]

In a simple two-group case where group 1 is committed and group 2 optimally coordinates internally, neither of these can happen, so inertia cannot be *ex post* excessive. In Figure 2, suppose that group 1 has irreversibly adopted (say) A. To be adopted by group 2, B must be substantially better: $u_B^2(n_2) > u_A^2(N)$, or equivalently $u_B^2(n_2) - u_A^2(n_2) > u_A^2(N) - u_A^2(n_2)$. That is, B's quality or price advantage (assessed by group 2) must outweigh the additional network benefit of compatibility with group 1 (assessed by group 2 when it adopts A). Of course, there is inertia: if group 2 values compatibility with group 1, B will fail unless it is much better than A. But to maximize total surplus *ex post*, group 2 should adopt B only if $u_B^2(n_2) > u_A^2(N) + [u_A^1(N) - u_A^1(n_1)]$. Group 2 internalizes only part of the social benefit of inter-group compatibility, and is thus too ready to strand group 1. Far from excess inertia, this model displays ex post "excess

[46] Farrell and Saloner (1987) analogize the first problem to horses tied to one another who will not wander far or fast, because none can move independently and staying still is more focal than moving in a particular direction at a particular speed. They analogize the second problem to penguins, wishing to dive for fish but concerned that the first one in is most vulnerable to predators.

momentum."[47]

This result instructively contradicts the popular intuition that inertia is obviously *ex post* excessive. But with more than two groups, *ex post* excess inertia may well occur, because optimal coordination among *ex post* adopters may well fail due to coordination problems and/or free-riding. To see this, return to the sequential adoption model of Farrell and Saloner (1985). Adopters $1, 2, ..., K$ arrive in sequence and, on arrival, irreversibly choose to adopt A or B. Because of idiosyncratic preferences or relative technological progress over time, adopters have different preferences between A and B. There are network effects: adopter i gets payoff $u^i_z(x_z)$, where x_z is the total number of adopters on his network $z = A, B$.

Arthur (1989) simplified this framework by assuming that an adopter gets network benefits only from previous adoptions, not future ones; thus adopters need not form expectations about the future. He showed that a technology favored by enough *early* adopters can become permanently locked in. If the relative network sizes ever become lopsided enough to outweigh the strongest idiosyncratic preferences, all subsequent adopters follow suit, because none wants to lead a new bandwagon, even if he knew that all future adopters would join it. There is a free-rider problem in overcoming an installed-base lead. Thus suppose that network effects make $x = 2$ much more valuable than $x = 1$, and that most adopters prefer B, but that by chance the first adopter prefers, and adopts, A. Adopter 2, then, who prefers B, must compare $u^2_B(1)$ against $u^2_A(2)$. He may adopt A only because $x = 1$ is so undesirable, in which case he and all subsequent adopters would pick A; while if he chose B, then other B-lovers would be happy choosing B thereafter.[48] This is extreme, but getting a new network up to critical mass can generally be costly for the pioneer, harmful to the installed base, but valuable to those who arrive later.

Arthur's assumption that adopters do not care about future adoptions seems to fit learning-by-doing with spillovers rather than most network effects, but we can usefully re-formulate it. Adopters more generally get the present value of a flow of network benefits, where the flow is increasing in adoptions to date. Then if adopter 2 adopts B and others follow, his sacrifice of network benefits is only temporary.

In this broader framework, Arthur's model assumes that adopters are infinitely impatient, thus both ignoring coordination problems and exagger-

[47]Farrell and Saloner (1986b) phrased this result in terms of "unique equilibrium" because they did not assume that each group optimally coordinates. Ellison and Fudenberg (2000) use essentially this model with optimal coordination to argue that there may be excessive innovation. If early adopters (group 1 here) would switch *ex post* to retain compatibility with group 2, group 2 is clearly again too willing to choose B. See also Shy (1996) and Witt (1997).

[48]This is similar to the "informational herding" literature: see e.g. Banerjee (1992), Scharfstein and Stein (1990), Ellison and Fudenberg (1993, 1995), Bikchandani, Hirshleifer and Welch (1992). Berndt, Pindyck and Azoulay (2003) argue that informational herding creates network effects in anti-ulcer drugs.

ating the free-rider problem. On the other hand, Farrell and Saloner (1986a) considered *ex ante* identical adopters with a finite discount rate. Adopters adopt immediately on arrival, and good B becomes available at date T. Specializing their model in the opposite direction from Arthur's, if identical adopters are infinitely patient *and* can optimally coordinate from any point on, the problem reduces to the two-group model outlined above in which *ex post* excess inertia cannot arise.

But the coordination problem re-emerges as soon as we depart from Arthur's infinite impatience. In particular, if previous history is the leading cue for coordination, then a patient small adopter 2 will compare $u_B^2(1)$ against $u_A^2(K)$,[49] so that an early lead would be even *more* powerful than Arthur's model suggests; it may be a self-fulfilling prophecy that a minority network will never grow. And if there are many contenders to displace the incumbent, adopters might expect splintering among those who abandon the incumbent (Kretschmer 2004). By the same logic, if everyone expects the new network to take over then it often will do so even if it is inefficient.

With identical adopters, the inductive logic of Farrell and Saloner (1985) suggests that the first adopter to arrive after T is pivotal. If he prefers that everyone forever stick to A, he can adopt A and thus make the next adopter feel all the more strongly the same way; similarly if he prefers that all from now on adopt B.[50] Because of the free-rider problem, the pivotal adopter may have too little incentive to adopt the new network, B; on the other hand, adopting B strands the installed base. As in 3.3.5 above, the net externality can run in either direction, so *ex post* excess inertia and excess momentum are both possible, even in unique equilibrium. If we eschew the long chain of backward induction and instead assume that the date-T adopter expects others' future choices to be unaffected by his own (he is small), then there are typically multiple equilibria and expectations determine the outcome, which can be biased in either direction. This would presumably also be the case if nobody knows which adopter is pivotal.

Farrell and Saloner (1986a) and Ostrovsky and Schwarz (2002) describe other models in which efficient coordination is hindered by delays before other adopters can follow an early mover's lead. Each is most easily described for two adopters. In Farrell and Saloner, each adopter has only occasional opportunities to adopt a new technology, so even if each adopts as soon as possible, adopting first entails a temporary loss of network benefits. If that is painful enough, neither adopter is willing to lead; the private cost may be either greater or less than the social cost. In Ostrovsky and Schwarz, each adopter chooses a "target" time to adopt, and if there were no noise, immediate adoption by all would be a Pareto-dominant equilibrium. But when actual adoption time is affected by (continuous) noise, Pareto-dominance is not

[49] This makes what may seem an unduly pessimistic assumption about later adopters' expectations if adopter 2 picks B. But that pessimistic assumption seems more natural if we are instead discussing adopter 3 after two A-adoptions.

[50] Thus his preference is evaluated assuming that all subsequent adopters follow his lead.

enough. Each adopter i can contemplate slightly delaying its adoption, by dt. If p_i is the probability that it will be the first to adopt, slight delay is privately desirable with probability p_i and then yields a gain of $[u_A^i(2)-u_B^i(1)]dt$; it is privately undesirable with probability $1-p_i$ and then yields a loss of $[u_B^i(2)-u_A^i(1)]dt$. Hence if $(1-p_i)[u_B^i(2)-u_A^i(1)] < p_i[u_A^i(2)-u_B^i(1)]$, or $p_i > r_i \equiv \frac{u_B^i(2)-u_A^i(1)}{u_B^i(2)-u_A^i(1)+u_A^i(2)-u_B^i(1)}$, it will prefer to delay slightly. Thus in any equilibrium with adoption by all, $p_i \leq r_i$ for all i. But $\sum p_i = 1$, so if $\sum r_i < 1$ then there is no equilibrium with adoption, even if all would gain ($u_B^i(2) > u_A^i(2)$ for all i) and there is only a little noise. However much each player expects others (collectively) to delay, he wants to delay slightly more.

Entry Our discussion of inertia also informs us about competitive entry of a product that is incompatible with an established network. Inertia implies that even if an entrant offers a better deal, network effects aside, to new adopters, they may (and perhaps should) stick to the installed base, assuming that the base itself won't move (perhaps because of individual switching costs). Incompatible entry is difficult, and Fudenberg and Tirole (2000) show that limit pricing can be powerful with network effects.

If new adopters optimally coordinate, this inertia is presumably because, for them, compatibility with the installed base outweighs the new product's advantages. As noted above, inertia can be *ex post* efficient given incompatibility,[51] although even *ex post* excess momentum (too-strong incentives for such entry) is possible. The point here is not whether incompatible entry is *too* hard *ex post*, given incompatibility and the installed base, but the fact that even efficient (indeed, even less-than-efficient) inertia can confer *ex post* market power on the established network.

But of course some incompatible innovation/entry succeeds. Inertia is sometimes overcome, and it is instructive to ask how. Of course, a product that is "ten times better" may simply outweigh inertia. But barriers can be lowered in other ways.

First, compatibility with the installed base eliminates the coordination and free-rider problems, and lowers individual switching costs; even partial compatibility through converters (see 3.8) can help. Similarly, multi-homing or double purchase (de Palma, Leruth and Regibeau 1999) mitigates pivotal adopters' losses of network benefits if they switch; Shapiro (1999) thus argues that exclusive dealing[52] by incumbents in network markets is especially threatening. Complementors can also multi-home, as when applications soft-

[51] Moreover, we saw that *ex post* excess inertia, blocking *ex post* efficient incompatible entry, is plausible when there are free-rider or coordination problems among adopters, and perhaps especially if expectations track history; Krugman (1991b) discusses the relationship between expectations and history. Since those problems may become more severe as the installed base grows, incompatible entrants may face "narrow windows" of opportunity (David, 1986).

[52] Broadly speaking this means agreements that make it hard for an entrant to thrive with small scale or limited scope.

ware providers "port" their programs from one operating system to another.

Rapid market growth makes the installed base less important relative to new adopters, and can thus mitigate pivotal adopters' transient losses of network benefits if they lead a switch (Farrell and Saloner 1986a); large players may both suffer less from such losses and be especially effective leaders of a bandwagon. When expectations otherwise focus on the incumbent, mechanisms such as consensus standards to help adopters coordinate on the best deal can also lower entry barriers. Finally, just as splintering among innovators tends to preserve the status quo (Kretschmer 2004), disarray and incompatibility in the installed base may open up opportunities.

As this last point suggests, successful static compatibility or standardization might retard (incompatible) innovation. Although the logic requires care—it is natural that the better the status quo, the less likely a good system is to engage in costly change—this might be an argument (in the spirit of maintaining biodiversity) against static standardization. But while marketwide compatibility may retard incompatible replacement of the compatible outcome, mix-and-match compatibility encourages component innovation (Langlois 1990).

3.5.2 Early Power

When there will be inertia—even *ex post* efficient inertia—early movers' choices determine later adoptions. Thus early movers might strategically or inadvertently commit to a standard that is bad for later adopters but won't be abandoned. We say there is *excess early power* if early movers adopt and are followed but this is ex ante inefficient: efficiency might demand instead that they defer to later adopters' preferences, or that they wait. That is, early adopters have excess power if their preferences weigh too heavily (relative to later adopters') in the collective choice of what is adopted.

Such an *ex ante* problem is sometimes called excess inertia, but we prefer to distinguish it more sharply from the *ex post* problem discussed above. They differ not only in timing, but in that ex post excess inertia concerns *later* adopters' choices, while ex ante excess early power concerns *early* adopters' choices. Excess early power does not imply ex post excess inertia: for instance, with two groups we saw that if group 2 optimally coordinates then there cannot be ex post excess inertia, but if inter-group network effects are strong and group 1 optimally coordinates, it has all the power. But the two concepts reflect the same force: the stronger ex post inertia will be, the more power early adopters have.

Arthur's model predicts excess early power; foresight complicates but does not fundamentally change the picture. Moving first gives commitment: early adopters are pivotal (early power), and the more they recognize that later adoptions will have to follow, the less sensitive early adopters will be to later preferences. Like inertia, early power can be efficient but can readily go too far: with strong network effects, long-run network technology choice can be determined by first-mover advantage and by historical small events.[53] With positive (not necessarily small) probability, almost all adopters choose A but total surplus would have been greater had almost all chosen B.[54]

Lock-in could go the other way, in which case foresight weakens early power: if group 2 finds adopting B a dominant strategy, while group 1 wants to adopt whatever it expects group 2 to adopt, then group 2 is pivotal.[55] But that requires network effects to be strong for group 1 but weak for group 2, so reverse lock-in seems likely to be rarer and weaker than forward lock-in. Thus Farrell and Saloner (1985) found that, given preferences, each player is better off moving earlier: this "New Hampshire Theorem" says that earlier adopters' preferences get more weight than later adopters' in the collective outcome,[56] which strongly suggests excess early power.[57]

In summary, early adopters have the strategic advantage: there is a reasonable presumption of excess early power at the adopter level. As we see in 3.7.2 below, however, this need not imply that early advantages confer sustained success when sponsors of competing standards compete using 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 any early lead in adoptions (whether strategic or accidental) will tend to expand rather than to dissipate. Network markets are "tippy": early instability and later lock-in.

To explore this, consider a continuum of identical adopters who only want to coordinate. There are three kinds of static pure-strategy Nash equilibria: all adopt A, all adopt B, and a continuum of splintered equilibria in which half adopt A and half adopt B (and all are indifferent). Now suppose 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 are

[53]Thus it can create a "butterfly effect:" a butterfly flapping its wings might cause a hurricane years later and thousands of miles away.

[54]In principle this might also arise if good A is worth more than B when each network is small but B is worth more than A when each network is large. As Liebowitz and Margolis (1994) observe, there is no obvious reason to expect that.

[55]Holmes (1999) shows how adopters who care less than others about network effects (relative to their preferences between products, or in his case locations) can lead a transition. He uses this in explaining the migration of the US cotton textile industry. Large groups that can successfully coordinate internally are thus prime candidates to be pivotal movers and get the best deals. Bresnahan (2001) explored this in the context of AOL's adoption of Internet Explorer during the Netscape-Microsoft browser war.

[56]The name arises 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.

[57]Excess late power (sometimes called *ex ante* excess momentum) is also possible, because the outcome depends only on ordinal preferences and not on their intensity.

unequal, those who can choose will adopt the majority product; this makes the half-and-half equilibrium unstable. The point carries over even with some horizontal product differentiation.[58]

Although sketchy, such dynamics suggest that when one infinitesimal player moves others will move too: re-equilibration by others strengthens instability. Such re-equilibration is central to indirect network effects, and may be a necessary condition for tipping: without it, Ellison and Fudenberg (2003) and Ellison, Fudenberg and Mobius (2004) show that there may be a plateau of non-tipped outcomes from which no player unilaterally wants to move, if buyers dislike (slightly) outnumbering sellers more than they like being in a bigger market.

Arthur (1989, 1990) and Arthur and Lane (1993) similarly find that if prices are fixed, and adoption decisions depend only on past adoptions (current shares of installed base), then one product or technology will come to dominate.[59]

3.5.4 Option Value of Waiting

We have seen that early adoption can freeze a technology choice and foreclose what would otherwise be later adopters' preferred choices. Above, we asked whether early adopters instead ought to defer to the known preferences of later adopters. When those preferences (and/or later costs) are not known early on, waiting can thus be efficient. Lock-in—even lock-in to a choice that's optimal given available information at the time—sacrifices social option value.

Just as future preferences are often under-weighted by market forces, option value will be. And institutions may be less apt to repair this: it is probably easier to acquire residual rights in one potential network with a clear future than to internalize the gains from waiting for something unpredictable. Whether or not the Dvorak keyboard is better than QWERTY, there clearly was a chance in 1888 that something better would later appear. How might incentives at that date incorporate this option value—what would persuade early generations of typists to wait, or to adopt diverse keyboards, *if* that was socially desirable in the long run? In principle the option value might be internalized by a century-long monopoly on typing (so that the monopoly could price the loss of option value into early adoptions), or by a futuristic patent on a range of alternative keyboards (so that Dr Dvorak's grandparents could subsidize waiting or diversity). Even if there had been many individual long-lived patents on particular keyboards, their proprietors would have faced a public-good problem in encouraging waiting. These institutions seem far

[58] With a finite number of adopters rather than a continuum, the same force prevents equal shares being an equilibrium at all. See e.g. Katz and Shapiro (1985, 1994). Echenique and Edlin (2002) show that strategic complementarities make mixed-strategy equilibria unstable, unless adopters have perverse beliefs about how shares will evolve.

[59] 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.

from reality. It might well *not* have been efficient for nineteenth-century typists to wait, or to use keyboards they didn't like, in order to preserve a more realistic option for a different design in 1940. But it is hard to think that the market gave a very good *test of whether or not* that would have been desirable.

Sometimes option value could be preserved by making later products compatible with early adoption. Section 3.8 below discusses incentives to do this, but clearly early adopters, or a sponsor of a product that they favor, may not want to ensure compatibility if they expect *ex post* inertia (excess or not) under incompatibility, as they gain from excess early power. Indeed, Choi (1994) and Choi and Thum (1998) confirm that pre-emption competition for the New Hampshire first-mover advantage can make adoption inefficiently fast when moving quickly can drive a bandwagon. Recall however that adoption may be too slow because of the externality or because early adoption risks coordination failure.

3.6 Price and Strategy for a Network Sponsor

Having discussed the demand side of network markets — adopters' choices given the offers they face — we turn to the supply side. This section primarily discusses a network monopoly, but most of the insights apply equally to a firm trying to establish its standard against a rival standard.

A sponsor seeking to establish its network has two generic strategies. First, it may focus selling effort on pivotal adopters, whose choices strongly affect others'. In particular, when a network involves different *classes of adopters* (for instance a credit card network that must be adopted by consumers and merchants) a sponsor can choose where to focus its marketing or price-cutting; and when there are different *adoption dates* a sponsor can choose (subject to commitment issues) when to do so. Second, a sponsor might seek to (visibly) *commit* to ensuring widespread adoption, or otherwise work on *expectations*.

3.6.1 Pricing to Different Groups: "Penetration Pricing"

First consider separate prices to two classes or groups of adopters with inter-group network effects.[60] These groups might be peers at different dates (early and late adopters), or two "sides" of a market. As Rochet and Tirole (2002, 2004) and Armstrong (2002) observe, such two-sided markets include credit cards, brokers, auctions, matchmakers, conferences, journals, computer platforms, and newspapers.

Suppose first that the sponsor simultaneously commits to both prices. Increased sales to one group raise the other group's demand: the inter-group "marginal" network effect. So in broadly Ramsey fashion the optimal price

[60]We consider only simple prices; Sundararajan (2003) discusses nonlinear pricing with network effects.

to group 1 will be lower, the more strongly group 2's demand responds to adoption by group 1 and the more profitable (endogenously) are sales to group 2, as well as the higher group 1's own demand elasticity (as usual).[61] Thus a single seller's optimal prices to the two groups may well be asymmetric; indeed, one side often pays zero or below cost.[62]

At an abstract level this is simply pricing with complementarities, as in Gillette's early strategy of giving away razors and making money on blades (Adams 1978); but here the complementarities are between different customers' adoption choices. If there is no single sponsor, implementing an optimal markup structure may require payments between sectors such as the credit card interchange fees discussed in 3.2; if that's hard to do well, it can encourage vertical integration.

With early and late groups the analysis is the same if the seller commits to a price path. For Ramsey-style reasons, penetration (low-then-high) pricing is privately (and can be socially) efficient in the usual case where early adopters are pivotal.

Finally, with early and late groups but no commitment, low-high pricing is even further encouraged. The seller will predictably set a second-period price higher than would be optimal ex ante, since ex post it will not take into account the effect on first-period adoption. Thus first-period adopters will expect a high future price, lowering first-period demand; and incompatible competition among sponsors will lower first-period prices in anticipation of the ex post rents. All these forces push towards bargain-then-ripoff penetration pricing, the reverse of Coasean dynamics.[63]

That commitment problem puts a sponsored network at a disadvantage against an "open" (competitively supplied) network product in the (relatively rare) case of reverse lock-in where second-period adopters are pivotal. A proprietary sponsor might then seek even costly forms of commitment such as (delayed) free licensing of a technology (Farrell and Gallini 1988, Economides 1996b). But sellers of an open product cannot recoup investment in

[61] As we noted in 3.3.2, there may be intra-group network effects (or congestion effects if the groups are different sides of a market). These affect the welfare economics, but for profit-maximizing pricing we can treat each group as a demand curve.

[62] See for instance Parker and Van Alstyne (2000), Schmalensee (2002), Rochet and Tirole (2004). As we saw in 3.3.2 above, first-best prices would be below marginal cost for both groups. Ramsey pricing looks qualitatively similar to profit-maximizing pricing because the problems are closely related.

[63] Cabral, Salant, and Woroch (1999) study 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: prices fall rather than rise over time, especially when there is complete information. Bensaid and Lesne (1996) find however that strong network effects remove the time-consistency Coase problem and cause optimal prices to increase over time. See also Mason (2000) and Choi (1994a). Radner and Sundararajan (2004) study a network monopolist's dynamic pricing problem when adopters expect each period's network size to be equal to last period's; they find extreme bargain-then-ripoff pricing (the monopolist prices at zero until the network reaches its desired size).

below-cost early prices, so a sponsored product has an advantage when (as is probably typical) overall adoption responds more sensitively to early prices than to (anticipated) later prices (Katz and Shapiro 1986a).

3.6.2 Single Monopoly Price

Above, we separated the two roles of p: each adopter viewed the price facing him in the ordinary way, and based his relevant expectations on the price facing the other (complementary) group. With switching costs, the ex ante and ex post prices are similarly separable when locked-in customers buy a distinct good such as service; otherwise they may have to be equal, as we discussed in 2.4. Similarly here prices to two sides of a market are presumably separable, but with two groups of peer adopters they may not be. In that case it is natural to suppress the two groups and simply study overall demand at the given price.

The "fulfilled-expectations demand curve" then matches each price p with those penetration levels x such that, when adopters expect penetration x, just x of them will adopt at price p: see e.g. Leibenstein (1950), Rohlfs (1974), Katz and Shapiro (1985), Economides (1996a). Such a demand curve is more elastic than each of the fixed-expectations curves of which it is built (Leibenstein 1950). Gabel (1991) suggests that Sony, Betamax's sponsor in VCRs, may have optimized against a less elastic (perhaps short-run) perceived demand curve because it did not anticipate video-rental network effects. Monopoly deadweight loss may be more severe with network effects: monopoly not only deters marginal adoption, but also lowers surplus of inframarginal adopters.[64]

Multiple equilibria in adoption at price p now show up as multiple intersections of the demand curve with a horizontal line at p. To pin down demand at p, one might rule out "unstable" equilibria (at which demand is upward-sloping); but if there is an unstable equilibrium, there are at least two stable equilibria. However one selects an adoption equilibrium for each p, there may well be discontinuous changes in behavior as a parameter such as cost varies continuously, as in catastrophe theory.[65] Even if a network product only gradually becomes cheaper or better over time, it may suddenly acquire "critical mass" and take off.[66]

[64] Farrell and Shapiro (1992) argue this in a linear example; but Lambertini and Orsini (2001), stressing network quality, reach different conclusions. One problem is that it is not clear what the demand curve "would be" without network effects. Perhaps more operationally, Rysman (2003) shows that, even if "competition" involves splintering, monopoly deadweight loss is worse in his calibrated model of the market for Yellow Pages.

[65] Indeed, if the rational-expectations demand curve has an upward-sloping portion, there is typically no everywhere-continuous selection of adoption equilibrium, even if there is everywhere a locally continuous selection.

[66] Rohlfs (2001), Farrell and Shapiro (1992), Economides and Himmelberg (1995), and Gandal (2000) suggest examples of sudden success that might reflect such tipping. Liebowitz and Margolis (1999) question that interpretation and argue that price and share dynamics in computer software seem inconsistent with tipping.

A strategic monopoly seller might persuade adopters to coordinate on the largest equilibrium x given p. If so, we say that the seller can "affect expectations" and pick any (x^e, p) such that x^e is an adoption equilibrium at price p. The next sub-section discusses some tactics for affecting expectations in this sense.

3.6.3 Commitment Strategies

Since demand depends on expectations, a network sponsor can gain from commitment to size, to inspire confidence and optimism. Commitment can address both the marginal and multiple-equilibrium underadoption problems identified in 3.3 above.

One form of commitment consists simply in selling products early on. Sellers boast about (even exaggerate) sales. To be a useful commitment, sales must be visible and irreversible, so this strategy makes most sense for durables. Network effects typically arise from use, not from mere possession, so dumping (e.g. free) units on the market may be discounted. The most effective sales are to "influential" adopters whose adoption will boost others' by the most.

The bluntest early-sales strategy is of course *penetration pricing*, as discussed above. As we will see in 3.7 below, competition can induce penetration pricing as the form of competition for the market. When a monopoly engages in penetration pricing, however, it would seem to be leaving money on the table relative to convincing early buyers in some other fashion that the long-run network size will be large. Thus we focus here on means to commit to that.

To encourage early adoption, a seller would like to commit to selling more later than it will then wish to sell, a point made by Katz and Shapiro (1986a) and put in a broader framework by Segal (1999). This kind of commitment strategy can operate even when there is a single equilibrium; commitment shifts the equilibrium. We have already noted some tactics such as second-sourcing that might help such a commitment. One might model commitment in a "reduced-form" way through assumptions about a sponsor's strategic variable. Rather than just setting a price, a sponsor might seek to commit to quantities sold or to the utility it will give each (type of) adopter: Armstrong (2002) shows how duopoly offers of customer utility levels yield different results than price-setting.

Reputation and general market credibility can help communicate commitment or boost expectations. Another commitment strategy is to open a standard to guarantee competitive future behavior, increasing early adopters' expectations of long-run network size. And integration with complementors might visibly improve incentives for supply of complements, as well as facilitate Ramsey-style cross-pricing.

When there are multiple equilibria, some of the same commitment tactics can help ensure a more favorable equilibrium. Rohlfs (2001) develops a model of irreversible adoption by many small buyers that involves dynamics at two levels. First, at any time buyers adopt if they want to do so given prices and given the current installed base, but they lack foresight and the adoption-equilibrium selection is thus pessimistic: there may be other equilibria with more adoption. In the second kind of dynamics, sponsors try to push the market past critical mass and generate positive feedback. For instance, a sponsor may dump enough units on the market to enter the basin of attraction of a preferred equilibrium.

In addition to the use of equilibrium-path price discrimination (penetration pricing), out-of-equilibrium (discriminatory) offers can eliminate an equilibrium that the seller dislikes, as we discuss next and as Segal and Whinston (2000) explored in the context of exclusive dealing. As that case illustrates, these equilibrium-selection tactics can work against buyers when networks compete, whereas in the case of a single network both seller and buyers prefer an equilibrium with more adoption.[67]

3.6.4 Contingent Contracts

Commitment through contracts could in principle overcome the coordination problem, as Dybvig and Spatt (1983) noted. Suppose a 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 $p' < u(n_i)$." Each buyer should accept this contract whatever he expects other buyers to do. Of course, p' may have to be (perhaps far) below cost, so the seller will make a loss if some buyers reject the offer. But in principle success depends only on buyers' individual rationality, not on their coordinating.

Likewise, the theory suggests, a contingent contract can profitably attract buyers away from coordination on the "wrong" network. Thus, suppose that buyers expect one another to adopt A, and that $u_B(n_i) - c_B < u_A(N) - p_A < u_B(N) - c_B$. Seller B offers the contract: "If x of you buy B, the price will be $u_B(x) - u_A(N) + p_A - k$." 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, as we noted in the previous sub-section, such a contract may be inefficiently successful: Segal (1999) and Jullien (2001) show that, because adoption of B imposes a negative externality on those who continue to buy A, there will be excessive adoption of B even if initial expectations favor A, when B (but not A) can offer public flexible pricing under complete information. But Park (2004) applies mechanism-design methods and finds that such contingent inducement schemes (and a range of other schemes) will induce less than efficient adoption when the seller has incomplete information about adopters' tastes.

[67]The reason is that one player's adoption of network A hurts—relative to the alternative—those who adopt B; thus in Segal's (1999) terms there is a negative externality on non-traders, leading to conflict at equilibrium when offers are public (full commitment by the seller).

It is not surprising that some flexible contracting can in theory solve coordination problems.[68] At the level of cooperative game theory, network effects are like ordinary economies of scale: in each case a coalition consisting of a seller and x buyers achieves more surplus per buyer as x increases. Indeed, Sutton's (1998, chapters 14.2 and 15.2) models of network effects and learning effects are formally identical. Since simple contracts often enable efficient competition with economies of scale (even dynamically if contestability holds), *some* contracts would in principle do so with network effects.[69]

Despite the theoretical appeal of contingent contracts, they do not seem the norm in network markets.[70] Implementation approaches may differ depending on whether adopters make a one-time purchase or continue to buy in order to use the network. Broadly, a seller might either charge low prices and later collect top-up fees if the network succeeds, or charge prices consonant with a successful network, promising refunds if the network falls short.

Very low, especially negative, prices may be problematic, as we discussed in section 2, and the nuisance adopter issue is arguably worse here because network benefits normally hinge on the use, not just possession, of the good. Especially if A is well established, this can make users' opportunity costs of adopting B large and hard to observe. Thus contingent contracts might work better against the single-network chicken-and-egg problem than to help an entrant displace an established network rival. But, when adopters will continue to trade with the seller over time, penetration pricing can become contingent pricing;[71] one version is usage-based pricing.[72]

The other broad implementation strategy of refund promises might not be believed, either because a nascent B-supplier would lack funds for such a large, non-diversifiable risk, or because buyers would suspect fine print in the contract.

While cost-side economies of scale often do not raise the coordination issues that we argue are central in network effects, this is not a fact of technology and preferences: it hinges on the contracts used. Thus contract theory should play more role in the study of network effects than it has hitherto.

---

[68] Thum (1994) also considers how contract form affects efficiency.

[69] One could also reach the same optimistic view via the Coase Theorem.

[70] Arguably this suggests either that there is no problem to be solved, or that (as we suspect) the contracts are problematic. See also Innes and Sexton (1994) and Haruvy and Prasad (2001).

[71] Another view of penetration pricing with one-time purchases is that it is an attempt at contingent pricing but sacrifices part of the surplus from early adopters: they "ought to" see that the network will succeed and hence be willing to pay a lot, but they don't.

[72] Oren and Smith (1981) and Rohlfs (2001). 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. See also Carter and Wright (1999).

3.7 Pricing of Competing Networks

In incompatible competition firms vie to control expectations. Competition will focus on pivotal customers; these are often early adopters—as with switching costs, where competition is largely for early purchases. Central questions are whether more efficient firms reliably win and whether profits reflect only their efficiency advantage.

3.7.1 Competition with Cost/Quality Differences

Consider incompatible competition with purely vertical differentiation: either a cost difference or a quality difference valued equally by all consumers. First we treat efficiency advantages as fixed over time; in 3.7.2 we allow them to vary. Expectations may respond in various ways to quality and price differences: for instance they may track surplus, track quality, track past success, or stubbornly favor one firm.[73]

We say expectations *track surplus* if each buyer expects all others to buy the product that, network effects held constant, offers the most surplus. For instance, suppose firms set prices just once and then there is a sequence of adoption choices by small cohorts. If adopters have similar preferences (agree on which product offers them more surplus if all adopt it), one might expect adoption of that product.[74] Price competition then works just as it would if the products were compatible. The efficient product wins, and (with non-drastic efficiency differences) consumers get the same surplus as they would if the second-best product were offered at average cost and adopted by all. Consumers capture the network effect and any economies of scale. Quality competition also is therefore just as under compatibility.[75]

But this changes dramatically if instead expectations *track quality*. Although this is a static model, this assumption can be motivated because, as Katz and Shapiro (1992) showed, this is the equilibrium if sponsors can adjust prices in response to adoption dynamics: suppose for instance that A has higher quality (or lower costs), and that this outweighs the network gain from adoption by a single additional cohort. Then, A will not fail through a bandwagon effect that starts because a *few* buyers adopt B instead. Rather, such a loss will lead A's sponsor to cut its price to subsequent adopters: it can profitably do what it takes to win, even coming from a bit behind in installed base.[76] So each adopter will recognize that even if he and his cohort adopt product B, product A will still win the rest of the market. Since no buyer

[73] These terms are from Farrell and Katz (1998).

[74] As we saw in 3.4.4, this is the unique subgame-perfect equilibrium. As we argued there, this may not be conclusive; but it is one plausible expection.

[75] Baake and Boom (2001), and Bental and Spiegel (1995) discuss static competition with network effects and quality differentiation when consumers' willingness to pay for quality varies.

[76] Therefore B will not attempt penetration pricing: there is no follow-on gain to winning a cohort or two. See Fudenberg et al. (1983) on races without leapfrogging.

is pivotal, the price to any buyer (or cohort) should not affect expectations. So rational expectations will *track quality*—focus on the network with higher quality (or lower costs)—and ignore any period's prices.

In this case, if A has higher quality it wins current sales if:[77] $u_A(N)-p_A \geq u_B(1)-c_B$, or $p_A - c_A \leq [u_A(N)-u_B(N)]+[u_B(N)-u_B(1)]-[c_A-c_B]$. Its profit is equal to its actual (cost and/or quality) advantage plus the network effect. If A visibly *could* make consumers a significantly better offer than can B, it need not *actually* match B's offer! Consumers would get more surplus if they all adopted the losing network B priced at cost.[78]

Of course, when such lucrative expectations track quality, firms will compete intensely on quality. Consumers gain from additional quality created by the second highest-quality firm.[79] The network effect accrues to the winner, and/or is dissipated in quality competition, which can therefore be socially excessive.

Worse, other factors might make consumers expect a product to win the market even after (out of equilibrium) losing a round or two — making expectations stubbornly unresponsive to price or performance. For instance, this logic would focus expectations on a firm that 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 roadmap for future products, control of a key complement, control of formal standards efforts, or marketing activity. As we saw, a seller thus favored by expectations can extract profits commensurate with the network effects, and may thus profitably control the market even with an inferior product or offering — provided, crucially, that its inferiority does not loosen its control of expectations. Such dysfunctional patterns of expectations may be most likely where adopters have dissimilar preferences, hindering attempts (e.g. through talk) to coordinate better.

When expectations thus *stubbornly* favor one firm, it has monopoly-like incentives for quality improvement. Its rivals cannot gain from ordinary innovation. But if B's quality improves so much that each user will adopt B no matter what he expects others to do, then adopters should now give B the benefit of expectations. Thus A's rivals have strong incentives for dramatic innovation (Grove's "ten times better").

Thus these models suggest that quality competition can produce stronger incentives for innovation than monopoly (even inefficiently strong incentives),

[77]We assume that each adopter is of size 1 and that a losing seller is willing to price down to cost.

[78]This is an instance of the principle that pivotal adopters get the surplus: when there are no such buyers, firms can keep the surplus. (Raskovich 2001 argues, on the other hand, that pivotal buyers find themselves saddled with the responsibility of ensuring that a good is actually provided.) In predatory pricing policy, Edlin (2002) discusses how a firm's ability to make a better offer can forestall the need to do so (to consumers' detriment).

[79]As always when competition gives no gross return to investment by a subsequent "loser", there can be equilibria in which only one firm invests. Thus details of the quality competition game may be important.

while expectations-dominant firms have incentives for incremental innovation and other firms have little incentive for other than breakthrough innovation.

If expectations track past market success, they reinforce installed base in giving past winners an advantage in future competition. This increases collective switching costs and accentuates the bargain-then-ripoff pattern of dynamic competition.

3.7.2 Competition with Cost/Quality Differences that Vary over Time

Now suppose that competing networks' efficiency advantages may shift over time. We revisit the inertia questions of 3.5 but now when competing networks are strategically priced. In doing so we address the scope for competitive entry (perhaps via penetration pricing) by a sponsored network product that must come from behind in network size and hence (often) in static efficiency, but that might become more efficient than an incumbent if widely adopted.

As we saw in 3.5, if early efficiency advantages determine offers to the pivotal early adopters, then a technology with an early lead will beat a technology that will (or may) be better later. This is the New Hampshire Theorem: early power for any given prices. In particular, if each network is competitively supplied, there is excess early power: a bias toward the one that early adopters prefer.

Now suppose instead that network "sponsors" compete for early adopters through penetration pricing. We describe how competitive penetration pricing can yield efficient adoption choices, and how biases can arise in either direction, but probably still with a tendency to excess early power.

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). Then A would price as low as $a_1 - (r - a_2)$ to win first-period sales, while B would go down to $b_1 - (r - b_2)$. Consequently, second-period efficiencies feed through efficiently 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 model, collective technology choice is efficient, and the pivotal (first-period) adopters get the benefit of competition.[80]

How robust is this optimistic result? Second-period efficiency can feed through *more* strongly than is efficient into first-period penetration pricing.

[80] Welfare may still be lower than under compatibility if different products would then be adopted in different periods, although firms have an incentive to achieve compatibility in that case (Katz and Shapiro 1986b; see 3.8 below).

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;[81] similarly B expects a second-period prize of $a_2 - b_2 + \beta$ for winning the first period. 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 excess late power[82] despite the excess early power for any given prices: strategic pricing here *reverses* the adoption-level bias.

Or feed-through can be *weaker* than is efficient. There is *no* feed-through when both standards are unsponsored (firms cannot later capture gains from establishing a product). Uncertainty and capital market imperfections can weaken feed-through.[83] Feed-through is also inefficient if first-period competition is not entirely through better offers but consists of rent-seeking through unproductive marketing. Feed-through can work efficiently even if consumers do not know why they are getting good first-period offers, or do not know the extent of gouging, provided the latter is symmetric. But, as we saw in Section 2, bargain-then-ripoff competition can cause inefficiencies.

As Katz and Shapiro (1986a) also noted, when one product is sponsored but its rival is not, feed-through is *asymmetric*, biasing the outcome toward the sponsored product. And, as Farrell and Katz (2001) note, feed-through is also asymmetric if A would stay in the market for the second period after losing the first, but B would exit if it lost the first round.[84]

To summarize, at given prices, network effects cause pivotal adopters' preferences to be over-weighted; since early adopters are often pivotal, products that appeal to them fare better than products that appeal comparably to later adopters. That is, there is typically excess early power for any given prices. But relative efficiencies in serving non-pivotal adopters may feed through into prices to pivotal adopters, and thus into the outcome. This feed-through can be zero (as with unsponsored products), weak, correct (as

[81]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.

[82]This is why Katz and Shapiro (1986a) find excess late power (or "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. The ("new") firm with the second-period advantage certainly would win second-period sales if it won first-period sales; but the other firm with the second-period disadvantage might not. The "old" firm would like to commit to doing so, in order to offer first-period customers a full network, but cannot.

[83]Feed-through will be weakened (as in switching-cost markets) 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 because negative prices attract worthless demand).

[84]Then, A's second-period prize for winning the first period is $r - a_2$, but B's is only $\min[r - b_2, a_2 - b_2 + \beta]$. Thus if $r > a_2 + \beta$, feedthrough is asymmetric and A wins both periods if and only if $a_1 - [r - a_2] \leq b_1 - [a_2 - b_2 + \beta]$, or $a_1 + a_2 \leq b_1 + b_2 + [r - a_2 - \beta]$. The last term in brackets is a bias toward the firm with a reputation for persistence.

in the model above where first-round losers exit), or excessive (as in Katz and Shapiro 1986a and Jullien 2001). Nevertheless, in general we think feed-through seems likely to be too weak, even if buyers optimally coordinate: the arguments for optimal or excessive feed-through put a lot of weight on firms' ability to predict future quasi-rents and incorporate them into today's pricing. Perhaps more importantly, however, feed-through can be asymmetric for reasons unrelated to the qualities of the competing products, and the asymmetry probably tends to favor established or sponsored products over nascent or unsponsored ones.

Thus entry by an incompatible product is often hard, and may well be *too* hard even *given* the incumbent's installed base and *given* incompatibility. Switching costs and network effects can work in tandem to discourage incompatible entry: switching costs discourage large-scale entry (which would require the installed base to switch) while network effects discourage gradual, small-scale entry (offering a small network at first).

A Switching-Cost Analogy The models above have close switching-cost analogies, although the switching-cost literature has not stressed efficiency differences between firms. With costs as described above and no network effects or quality differences 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 he is correct about second-period pricing (for instance, if his 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 with lower life-cycle costs makes the sale, as efficiency requires. This is the switching-cost analogy to the model with exit above.[85]

But if second-period prices are instead 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 and know firms' second-period costs, then the buyer chooses 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.

3.7.3 Static Competition when Consumers' Preferences Differ

Without network effects, or with compatibility, horizontal differentiation has several effects. First, tipping is unlikely: a variety of products make sales. Second, prices reflect each firm's marginal cost and its market power due to

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

the horizontal differentiation: in a Hotelling model, for instance, the level of transport costs will affect prices. Third, if any seller modestly improves its product, it yields modestly higher share and profits.

With strong network effects and incompatibility, all these lessons change. Buyers want to coordinate and all adopt a single network, though they disagree on which one. If they will succeed in doing so, and if their collective choice is responsive to changes in quality or price, then firms are competing for the market, which blunts horizontal differentiation. Thus, strong proprietary *network effects can sharpen price competition* when expectations are up for grabs and will track surplus;[86] Donanoglu and Grzybowski (2004) contrast this with competition-softening switching costs. Product improvement by the leader does not change market shares; nor does marginal product improvement by other firms. If price reflects cost, it will be average cost,[87] the relevant cost when competition is for the market and all-at-once.

When differentiation is stronger, or network effects weaker, niche minority products such as Apple can survive. Multiple products can also survive if network effects are primarily within two or more groups, in which case the market is segmented. 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, as speakers of "small" human languages are finding and as Gabel (1987) argues was the case for Betamax.

Thus if expectations respond sharply to relative surplus, network effects can sharpen competition.

3.7.4 Dynamic Competition when Consumers' Preferences Differ

Just as excess early power at fixed prices need not imply excess early power when firms compete in penetration pricing, tipping at given prices might not imply tipping when sponsors price to build or exploit market share. If one network gets ahead, will its sponsor raise price to exploit that lead (thus dissipating it), or will it keep price low and come to dominate the market? Arthur and Rusczcynski (1992) studied this question when firms set prices in a many-period dynamic game; Hanson (1983) considered a similar model.[88] In stochastic duopoly they find that if firms have high discount rates, a large firm tends to lose share by pricing high for near-term profit. But if firms have lower discount rates, a large firm sets low prices to reinforce its dominant

[86] 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.

[87] That is, the loser's price will be its average cost. The winner's price may be higher, as we have shown.

[88] See also Keilbach and Posch (200x).

position.[89]

Thus with strong network effects, tipping or unstable (positive feedback) dynamics is likely with given prices (including the case of unsponsored standards and constant costs); moreover can readily—though it need not—also arise where firms strategically set prices. Recall by contrast (see 2.7) that, with switching costs, repeated sales of a single good, and no price discrimination, firms with large installed bases normally price high and market dynamics thus tend to be stable; small-scale entry is relatively easy.

3.8 Endogenous Network Effects: Choosing How to Compete

Incompatibility of competing products is sometimes inevitable, but is often a choice. Why would a firm prefer one form of competition over another, and who gets to choose?

3.8.1 Efficiency Effects

Incompatibility has some obvious inefficiencies. Network benefits are lost if the market splinters, as it may well do if network effects are weak or if adopters choose simultaneously or in ignorance; network effects are also lost if the market changes its collective mind. If, on the other hand, the market cleanly tips, it worsens matching of products to consumers when tastes differ or if the market tips the wrong way. When networks' future relative advantages are uncertain, compatibility makes switching easier (whether or not inertia is efficient given incompatibility) and thus preserves option value and reduces adopters' incentives to wait and see which network wins or to adopt hastily and pre-empt.

Compatibility can also enable mix-and-match of complements. When the best hardware and the best software may not come from the same family, there is a direct mix-and-match efficiency gain from compatibility.

Efficiency costs of compatibility include direct costs of design constraints or of adapters to ensure compatibility; design constraints may be severe when a standard is controlled by a slow-moving consensus process. Proprietary control of a standard can encourage investment in development or in penetration pricing.

When firms do not compete, or when competition is equally fierce either way, efficiency effects should govern: firms will internalize efficiency advantages of compatibility choices. But competitive effects modify this, and can readily reverse it.

[89] Dosi, Ermoliev and Kaniowski (1994) find that market sharing can occur if firms adjust prices in response to market shares according to an exogenous non-optimal rule.

3.8.2 Competitive Effects

The first competitive effect is *leveling*: compatibility neutralizes the competitive advantage of one firm having a larger installed base or being better at attracting expectations. When firm 1 is larger than firm 2, so $x_1 > x_2$, compatibility boosts the value of firm 1's product from $u(x_1)$ to $u(x_1 + x_2)$, and firm 2's product from $u(x_2)$ to $u(x_1+x_2)$. Since a firm's profit is increasing in the value of its own product and decreasing in that of its rival, compatibility helps the large firm less and hurts it more than it helps or hurts the small firm *if* we can take the (expected) sizes x_1 and x_2 as broadly given. So a big firm, or one that is exogenously expected to be big, is more apt to resist compatibility, especially if competition between the firms is strong.[90]

Thus the dominant Bell system declined to interconnect with upstart independents in the early post-patent years of telephone competition in the US, and Faulhaber (2002, 2003) describes AOL's failure to interlink with rivals' instant messaging systems. Borenstein (2003) similarly argues that interline agreements between airlines, which let customers buy discount tickets with outbound and return on different airlines, help smaller airlines much more than larger ones; interlining has declined over time. Bresnahan (2003) describes how Word Perfect sought compatibility with the previously dominant WordStar, but then fought compatibility with its challengers.

Second is the *un-differentiating effect*. As in 3.7.3, when tipping is likely and size is (or expectations are) completely up for grabs, incompatibility can neutralize ordinary horizontal differentiation that would soften price competition in compatible competition. Even when it is less efficient, incompatible competition can then be sharper. But when tipping is unlikely, incompatibility can *create* horizontal differentiation (segment the market), as in switching-cost markets; Augereau, Greenstein and Rysman (2004) find that when ISPs chose between incompatible 56kbps modems, there was less compatibility than random choice would imply in each local market, which they attribute to ISPs' desire for horizontal differentiation.

Thus firms' incentives will depend on the likelihood of tipping and on whether expectations are largely exogenous or are symmetrically competed for. Real-world frictions, including switching costs, limit short-run shifts of customers (or expectations). If simple network models understate such frictions, they may thus overestimate the strength of incompatible competition.

Third, if each side has proprietary complements that remain fixed independent of scale, and compatibility enables mix and match, duopoly models suggest that firms' private gains from compatibility exceed the social gains, but this is less clear with more than two firms (see section 2.8.4). We digress briefly here to discuss the relationship between these mix-and-match models

[90] See for instance Katz and Shapiro (1985), de Palma and Leruth (1996), Economides and Flyer (1997), Cremer, Rey and Tirole (2000), and Malueg and Schwartz (2002). Belleflamme (1998) explores how the leveling effect varies with the number of firms and with the form (e.g. Cournot vs Bertrand) of competition. It may be particularly unfortunate if large players resist compatibility, since they tend to be best at leading bandwagons.

and indirect network effects.

Indirect Network Effects and Mix-and-Match Both indirect network effects and the mix-and-match literature discussed in 2.8.4 above study modularity (mix-and-match) versus proprietary complements in a systems market, but the two literatures are surprisingly hard to relate; we note some key differences, but future research should develop a more unified understanding.

When more customers buy "hardware" of type A, the demand for A-compatible "software" increases, so there is more profit to be made from providing such software if entry does not dissipate that profit. The mix-and-match literature, like the bundling literature (e.g. Nalebuff 2000), allows for this profit increase to be captured by the A-hardware provider through vertical integration. It then studies pricing and profits when this fact *does not induce additional entry* into A-compatible software.

In contrast, as we discussed in 3.1, the indirect network effect literature assumes that when more A-hardware is sold, the boost in A-software demand *does induce additional (re-equilibrating) software entry*, making A's hardware more attractive to customers and thus indirectly increasing hardware profits. But a boost in software profits is not part of this calculation, both because entry dissipates software profits and because most models assume there is no integration.

We also note that with indirect network effects, tipping at the hardware level increases software variety while reducing hardware variety.[91]

3.8.3 Institutions and rules: Who chooses?

If participants disagree on compatibility, who chooses? This question arises at several levels. We pose it primarily as a tussle among competing vendors with different preferences over how to compete. Another version of the question pits one vertical layer against another: often customers against vendors. A third version concerns the various means to achieve network benefits. Finally, there may be (as in TV) compatibility domestically but not internationally.

a. Among horizontal competitors Sometimes side payments can be made smoothly enough that the outcome is the one that maximizes *joint* profits. If side payments are fixed or one-shot, efficiency effects and the ferocity/softness of competition will drive the joint decision. And if firms can charge one another running royalties for compatibility, that may itself soften compatible competition. In telecommunications, interconnection (compatibility) is largely compulsory but charges for interconnection are common;

[91] When indirect network effects are proprietary (mixing and matching is impossible), tipping at the hardware level tends to *improve* the match between customers' software tastes and the software varieties endogenously provided, by increasing the size of the winning hardware platform's market (though tipping worsens hardware matches).

Ennis (2002) shows that the curvature of the network-benefit function can determine equilibrium payments, while Hermalin and Katz (2003) show how efficient carrier-to-carrier pricing depends on demand elasticities. Brennan (1997) and Laffont, Rey and Tirole (1998a,b) ask whether competing firms can use such charges to support monopoly outcomes as noncooperative equilibria. Similar concerns may arise if firms agree to include one another's intellectual property in a consensus standard or a patent pool, as Gilbert (2002) stresses.[92] But these strategems might be hard to distinguish in practice from side payments needed to encourage efficient compatibility.

In other cases firms choose how to compete noncooperatively without smooth side payments. As above, any firm wants to offer its customers bigger network benefits, and wants its rival's customers to get smaller network benefits. Thus each firm would like to *offer* a one-way converter that gives its customers the network benefits of compatibility with its rivals' customers; but would like to *block* converters in the other direction.[93] In a noncooperative framework, then, if any firm can block such a one-way converter (e.g. through intellectual property or by secretly or frequently changing an interface), incompatibility results. But if any firm can unilaterally offer a one-way converter, compatibility results.

One can then study incentives for two-way compatibility by thinking of converters in the two directions as inseparably bundled. There are three possibilities: neither side wants compatibility; both do; or they disagree.[94] When the firms disagree, incompatibility results if the firm who dislikes compatibility (typically the larger or expectations-dominant player) can prevent it, perhaps through intellectual property or through secrecy or frequent changes of interface. MacKie-Mason and Netz (2004) explore micro-analytics and institutions of such strategies. On the other hand, compatibility results if it is easier to imitate than to exclude, as Gabel (1991) argues it was for auto parts.

With more than two firms, compatible coalitions may compete against incompatible rivals.[95] Extending Katz and Shapiro (1985), Cremer, Rey, and Tirole (2000) describe a dominant firm's incentive for targeted (at one smaller rival) degradation of interconnection even if it has no incentive for uniform degradation. But Malueg and Schwartz (2002) observe that a com-

---

[92] Firms might also sustain price collusion by threatening to withdraw cooperation on compatibility.

[93] See Manenti and Somma (2002). Adams (1978) recounts how Gillette and others fought this battle of one-way converters in the razor/blade market. Of course, like most statements about oligopoly, these arguments are subject to subtle commitment effects.

[94] Besen and Farrell (1994) analyze compatibility choice in these terms. Farrell and Saloner (1992) analyzed effects of two-way converters, and also found that converters can reduce static efficiency; Choi (1996b, 1997b) finds that converters can block the transition to a new technology. See also David and Bunn (1987), Kristiansen (1998), and Baake and Boom (2001).

[95] Axelrod et al. (19xx), Economides and Flyer (1998), and Farrell and Shapiro (1993) also study coalitions in network markets with more than two players.

mitment to compatible competition may attract users and deter degradation; Stahl (1982), Dudey (1990), and Schulz and Stahl (1996) similarly discuss incentives to locate near competitors. Cusumano et al. (1992) suggest that this was important in VHS' victory over Betamax.

b. Vertical locus of compatibility choice Network benefits can result from choices at various vertical layers (see section 3.3.2). The efficiency effects may broadly be the same, but competitive effects may differ according to the vertical layer at which compatibility happens. Many consensus standards organizations bring together participants from multiple layers, though few true end users attend. The literature's focus on competing interests is a simplification of the web of interests that results. In particular, end users often compete with one another less than do the vendors who sell to them, making it easier for end users than for vendors to agree on standards; but there are typically many end users, making it hard.

A value-chain layer with a single dominant provider may also be a relatively likely locus for standards. Thus for instance Intel has championed, even imposed, compatibility in some layers complementary to its dominant position. In favorable cases, a dominant firm has salutary incentives to influence complementary layers.

c. Means to network benefits One way to achieve network benefits is that all the players at one vertical layer of a value chain—perhaps vendors, perhaps end users—decide to adopt the same design. That in turn can happen through various mechanisms of coordination, including consensus agreements and sequential bandwagons, but also including tradition, authority, or the use of sunspot-like focal points. Another path to network benefits is the use of converters or adapters,[96] or the related multi-homing strategies such as learning a second language.[97]

d. International Trade Just as firms might choose incompatibility for strategic advantage, so too may nations pursuing domestic (especially producers') benefits at the expense of foreigners'. As in strategic trade with economies of scale, one strategy conscripts domestic consumers as a protected base to strengthen domestic firms in international competition: incompatibility may be a tool to do so, and Crane (1979) argues that this

[96]See David and Bunn (1990), Farrell and Saloner (1992), and Choi (1994, 1997). Because converters affect competition between otherwise incompatible networks, they may be subsidized or provided by sponsors of networks or may be independently supplied. Because network transitions are not first-best, strange effects can occur: for instance Choi shows that they can retard a transition.

[97]See de Palma et al. Multi-homing is also discussed in the context of two-sided markets by Rochet and Tirole (2004).

was why governments imposed incompatible standards in color television.[98] As with competing firms, Jensen and Thursby (1996) note that a country may prefer compatibility when its standard is behind, but will shift to preferring incompatibility if it wins. Gandal and Shy (2001) argue that countries will not choose standards autarky but may inefficiently form standardization unions that exclude some countries (as indeed happened in color TV).[99]

3.9 Network Effects and Policy

Policy intervention may encourage compatibility so as to avoid coordination failures or externalities. Compatibility is also important and problematic in competition policy.

3.9.1 Compatibility and Coordination Policy

If adopters are inefficiently splintering (failing to standardize) because of confusion, or are adopting too slowly because they fear coordination failures, a non-binding or indicative standard might help, as might simple discussion. As we noted above, the popularity of non-binding consensus standards suggests that this is a common situation. We defined "splintering" as an inefficiently incompatible equilibrium that is not coalition-proof; the solution accordingly requires leadership but not compulsion. Government may be a natural leader, as Besen and Johnson (1986) suggest in AM stereo, even though its comparative advantage in compulsion is not called for;[100] a firm that is dominant in a complement may also assert this role.

If failure to standardize is not a coordination problem, but stems from differences in preferences (horizontal differentiation), then an indicative standard will not help but a compulsory standard might, and government's compulsion may be needed. By driving on the wrong side of the road, a driver sacrifices his own network benefits from driving on the same side as others, but this may not discourage him enough. Administratively coordinated groups such as firms often impose compatibility internally, suggesting that they expect decentralization to yield too much incompatible variety. But in general (see 3.3.5) when network goods compete the net externality is ambiguous and relatively weak; moreover, compulsory standards can be wrongly chosen. And indeed firms by no means always impose internal standards.

[98] Farrell and Shapiro (1992), and Rohlfs (2001) discuss this in terms of network effects. Note also that US high-definition standards however contain many "options," which might threaten compatibility.

[99] Walz and Woeckener (2003) also find forces for inefficient incompatibility in trade policy. Kubota (1999) notes that transfer payments can make this less likely. Adams (1996), Choi, Lim and Yu (1999), Gandal (2002), Matutes and Regibeau (1996), and Klimenko (2003) also study trade policy with network effects.

[100] Development economists have studied "indicative planning"—government plans without compulsion.

Finally, as David (1996) suggested, the difficulty (even if efficient ex post) of incompatible transitions once a standard is established might suggest a public policy interest, along the lines of biodiversity, in postponing full *de facto* standardization, assuming that innovation is inevitably incompatible. But Cabral and Kretschmer (2004) find that in Arthur's (1989) model it is ambiguous whether public policy should seek to postpone or to accelerate de facto standardization.[101]

3.9.2 Compatibility and Competition Policy

As we have stressed, incompatible competition with network effects is often clumsy, and certainly differs from compatible competition. With (strong) network effects firms compete over sales to large groups of buyers rather than over single transactions. So, as with switching costs, some conventional cues to the state of competition may be misleading. For instance, comparing price with cost is a subtle matter. With switching costs, prices below costs *ex ante* and above costs *ex post* may simply reflect life-cycle competition. With network effects, large market shares and prices below cost may reflect competition for the market.

The fact that, with network effects, each buyer wants to conform with other buyers does not inevitably cause an overall competitive problem. Even if each buyer must follow the crowd, something drives the crowd's choice, and in favorable cases 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: competition for the market *can* work well.[102] Indeed, if expectations track surplus, incompatible competition can be even fiercer than compatible competition, by neutralizing horizontal differentiation (although it then worsens matching).

But although proprietary network effects need not cause competitive problems, they often do. Because incompatible entry is hard (whether or not the inertia is "excess") a firm that controls a standard may not be well disciplined by competition ex post even if buyers optimally coordinate subject to the installed base's reluctance to shift. Differences in preferences, which are ubiquitous when there is an installed base or other vested interests, make it less likely that buyers coordinate well through bandwagon mechanisms or through consensus standard-setting.

Coordination problems among adopters can weaken or misdirect competitive pressure. Entrants may be subject to splintering. If expectations

[101] Arthur (1988) might seem to suggest that subsidies or interventions to change the standard are the main tool of policy, but he 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.

[102] Kristiansen and Thum (1997) stress that network size is a public good in compatible competition.

favor incumbents, entry will be especially hard, and an incumbent may profitably control the market even after a more efficient rival does enter. This strengthens the prospects for recoupment of anticompetitive effects.

Even if coordination will be optimal *ex post*, there will be collective switching costs. Therefore, as we have stressed throughout, incompatible competition tends to feature fierce competition for pivotal adoptions and little competition for others. As we saw in section 2 and in 3.7 above, prospective profits from dominance do not always straightforwardly feed through into competition for dominance. More generally, competition for the market differs from, and is often less efficient than, competition in the market.[103]

There is a spectrum of views of such competition for the market. David (1985) and Arthur (e.g., 1989) seem highly pessimistic: long-run technology choice is inefficiently driven by accidental short-run small events. At the other extreme Liebowitz and Margolis (e.g., 1994) optimistically argue that neither coordination problems nor indirect network effects are a serious source of concern (while direct network externalities are relatively rare). Between these extremes, Bresnahan (2003) suggests that in the computer industry long periods of lock-in are punctuated by occasional "epochs" of competition for the market when barriers due to network effects and switching costs are much lower than usual because of a shift of the incumbent's standard or a strong independent complement.

All this suggests several ways in which network effects can affect good competition policy.

a. Recognizing the differences. Simply recognizing these differences can inform competition policy. Incompatibility often weakens or stops small-scale competition and entry; it may thus make predation and exclusion easier and more tempting than in compatible or non-network competition.

Likewise, policy may stop mergers that would create network- or expectations-dominant positions, whether because products are already incompatible or because they might become so: one of the competition authorities' concerns about the proposed MCIWorldcom-Sprint merger was that the combined firm would have so large a share in the internet backbone market that it might profitably deny efficient interconnection.[104]

b. Protecting fragile incompatible competition. Second, taking incompatibility as given, policy may protect and strengthen incompatible competition, which (we argued above) can work well but is fragile. In

[103] Demsetz (1968) is often cited on competition for the market, although the idea goes back to Chadwick (1859). Contestability (Baumol, Panzar and Willig (1988)) is closely related.

[104] See Cremer et al (2000), Dewatripont and Legros (2000), Ennis (2002), and Malueg and Schwartz (2002). Robinson (1998) describes how the Department of Justice addressed these concerns in the previous MCI-Worldcom merger.

particular, since incompatible competition can tend towards serial monopoly, it is important to protect entry and transitions.

Where there are indirect network effects, vertical integration may both promise direct efficiency gains and also raise competitive concerns since strong independent complementors will tend to be important potential entrants, as Bresnahan and Greenstein (1999) argue in the computer industry; the *Microsoft* cases, including the (abandoned) "divest Office" remedy echoed this logic.[105]

Exclusion and predatory pricing may be more feasible and tempting when there are strong proprietary network effects. 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 a rival-weakening incentive to do so.[106] 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 rules against predatory pricing in a network market are unlikely to be efficient. Below-cost pricing and short-run sacrifices to build market share may be legitimate competition for the market: see 3.6.–3.7 above, Cabral and Riordan (1997), and Farrell and Katz (2001). Large-scale (especially government) procurement may sensibly take into account the effect on future competition and eschew offers by sponsors of proprietary networks that are more attractive (e.g., cheaper) in the short run than competing open networks.[107]

Non-price predation 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). Gandal et al. (2003) found a significant effect of a preannouncement in DVDs. But Fisher (1991) argues that anticompetitive effects of preannouncements will be very hard to diagnose; see also Haan (2003).

Because incompatible competition works better when users coordinate their expectations well (in a way that tracks surplus), policy might sensibly seek to help them do so. In particular it could try to protect consensus standards-setting against unfair dealing by participants (or others). One

[105]The trial court in the Microsoft case approved a proposed remedy of breaking up Microsoft into an operating system company and one that would initially sell applications, including Microsoft's popular "Office" suite; the appeals court overturned this.

[106]Anton and Yao (1995) and Lemley and McGowan (1998) discuss antitrust concerns. Liebowitz and Margolis (1996) suggest that the concern is overblown. Katz and Shapiro (1999) discuss antitrust in software markets, where network effects are often important.

[107]This has been argued, for example, in debates about government procurement of information systems based on Linux versus those offered by Microsoft, when Microsoft offers substantial inducements such as free training, etc.

More generally procurement can be a very powerful public policy tool influencing compatibility and standard setting.

concern arises if a consensus standard is priced, often because it draws on intellectual property. Once a standard is established, inertia makes it possible to charge a price that exceeds the standard's inherent advantage over what would ex ante have been viable alternatives: if it is hard to coordinate a collective shift ex post then the intellectual property's owner can expropriate the network effects. Thus holdup can arise unless licensing is negotiated before momentum develops around the standard; anticipation of such problems exacerbates the delays due to vested interest. Standards organizations often try to require disclosure in advance and "reasonable and non-discriminatory" (RAND) licensing; but much of the harm from hold-up is borne downstream, so standards organizations (unless they actively involve end-users) have weak incentives to enforce rules for this problem. In particular, while other participants have complained about holdup behavior (for instance by Wang, Dell, Unocal and Rambus), problems could also arise if firms charge one another running royalties for compatibility, or agree to incorporate one another's intellectual property in a standard: see Gilbert (2002) and Laffont, Rey and Tirole (1998a,b).

At the same time, consensus standard-setting might itself raise competitive concerns. Agreeing on a common standard can eliminate inter-standard competition, presumably replacing it with compatible competition. Since the latter is often more efficient, this has not been a big concern, but it could be, especially if firms choose how to compete with an eye to competitive effects and with little end-user involvement. One safety valve is that private consensus standards are normally non-binding, though this is weak comfort if network effects are strong.

Exclusion of some rivals from a broad compatibility standard raises another concern, especially if incompatible rivals cannot really compete; many standards organizations specify that their agreed-on standards should be available on a nondiscriminatory basis.

c. Duties or rights of compatibility. When incompatible competition seems a poor substitute for compatible competition, policy may require compatibility. At its strongest, this policy might do so even if firms would all prefer incompatibility. Telecommunications policy broadly imposes interconnection (compatibility) among firms, although firms often charge one another for it; policy often also regulates such charges. The FCC partially extended this approach to instant messaging by imposing conditions on AOL's merger with Time Warner.[108]

A weaker pro-compatibility policy would let rivals insist on compatibility, for instance protecting a right of reverse engineering despite intellectual

[108] In the US the 1996 Telecommunications Act requires all telecommunications providers to interconnect, but allows competing local carriers to negotiate charges to do so (subject to default regulation and a public interest test). Sometimes firms agree not to charge for interconnection, a practice known as "bill and keep." In some other cases the charges are regulated.

property protection. Lemley and McGowan (1998a), Menell (2002), and Samuelson and Scotchmer (2002) discuss such rights.[109] Related to this are policies of open access to important standards, as in credit cards.

Still weaker steps in this direction would protect behavior that weakens (the effects of) incompatibility—such as converters and multi-homing. When network effects are indirect, protecting multi-homing means limiting exclusive dealing and encouraging "porting," as the FTC did in the *Silicon Graphics* merger. Shapiro (1999) argues that exclusive dealing by an incumbent can buttress the already high entry hurdles faced by incompatible entrants, making exclusivity especially worrying in network markets with proprietary standards.

d. Vertical Issues Economides and White (1994) observe that proprietary indirect network effects arise from vertical exclusivity (*e.g.*, exclusive dealing with complementors), whether by agreement or otherwise. Thus horizontal questions about competition with indirect network effects are also inherently vertical.

But there is a more purely vertical issue. Whereas above we asked how vertical exclusivity (incompatibility) affects horizontal competition, we can also ask whether the proprietor of a bottleneck at one layer might use its control of a standard or interface to leverage into a related layer. Proprietary interfaces can help a firm weaken or exclude independent complementors/competitors *if* it chooses; indeed it may be hard to commit to not doing so.[110] Thus the key question is whether there is an inefficient motive to do so.

The "one monopoly rent theorem" argues that dominant firms often do not want to export their dominance inefficiently, and indeed powerful firms often encourage competition in complements; this argument underlies the broadly laissez-faire vertical policy of modern antitrust.[111] But as Whinston (1990), Choi and Stefanadis (2001), and others have shown, the argument is not conclusive.[112] Much telecommunications policy, and "essential facility" ideas in antitrust, seek to protect competition that complements a monopoly platform.

[109] Farrell (1989, 1995) describes some arguments for weaker intellectual property protection in network industries. But Llobet and Manove (2003) argue that incumbents will build smaller networks if entrants may be able to share them, and recommend R&D subsidies rather than compatibility rights.

[110] See e.g. Henderson and Gawer (2003) and Farrell and Katz (2000).

[111] Farrell and Katz (2000) and Economides and Viard (2003) also discuss pricing of a monopolized "base good" and its complements.

[112] The "one monopoly rent theorem" points out that a monopolist in good X shares consumers' interests in efficient and competitive provision of a strict complement Y and so, in the simplest models, will not want to allow or engage in anticompetitive behavior in Y. Farrell and Weiser (2003) discuss the result and "exceptions" to it.