

Too expensive to meter: the influence of transaction costs in transportation and communication

BY DAVID LEVINSON¹ AND ANDREW ODLYZKO^{2,*}

¹*Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive Southeast, Minneapolis, MN 55455, USA*

²*Digital Technology Center, University of Minnesota, 499 Walter Library, 117 Pleasant Street Southeast, Minneapolis, MN 55455, USA*

Technology appears to be making fine-scale charging (as in tolls on roads that depend on time of day or even on current and anticipated levels of congestion) increasingly feasible. Such charging also appears to be increasingly desirable, as traffic on roads continues to grow and costs and public opposition limit new construction. Similar incentives towards fine-scale charging also appear to be operating in communications and other areas, such as electricity usage. Standard economic theory supports such measures and technology is being developed and deployed to implement them. But their spread is not very rapid and their prospects for the future are uncertain. This paper presents a collection of sketches, ranging from ancient history to very recent developments, that illustrate the costs that charging imposes. Some of those costs are explicit (in terms of the monetary costs to users and the costs of implementing the charging mechanisms). Others are implicit, such as the time or the mental processing costs of users. These argue that the case for fine-scale charging is not unambiguous and that in many cases such charging may lead to undesirable outcomes.

Keywords: transport; telecommunications; London Underground; London railways; collection costs; transaction costs

1. Introduction

Lewis L. Strauss, the Chairman of the US Atomic Energy Commission, is often remembered for a famous claim, made in the optimistic early days of nuclear power,

Our children will enjoy in their homes electrical energy too cheap to meter, It is not too much to expect that our children will know of great periodic regional famines in the world only as matters of history, will travel effortlessly over the seas and under them and through the air with a minimum of danger and at great speeds, and will experience a lifespan far longer than ours, as disease yields and man comes to understand what causes him to age.

(Strauss 1954)

* Author for correspondence (odlyzko@umn.edu).

One contribution of 16 to a Discussion Meeting Issue 'Networks: modelling and control'.

Strauss' expression 'too cheap to meter' has entered the lexicon as a catchphrase for technological promises that have gone unfulfilled. Electricity continues to be paid for roughly in proportion to usage. The cost of electricity is not so low, nor the cost of metering it so high, that the purveyors of electric power have chosen to forgo charging by use. The alternative, either 'free' service or a fixed price for unlimited (unmetered) use, has not come to pass in the electricity sector. Just the opposite appears to be happening. 'Smart meters' that charge different prices, either depending on time of day or, in even more sophisticated versions, depending on the level of electricity consumption in the system, are already deployed in industry and are becoming more widespread in residential use. In spite of continuing substantial progress in the electric power generation technology, overall costs are, if anything, increasing as fuel prices rise, and there is intense public opposition to building more power plants and transmission systems, as well as concerns about pollution, climate change, fuel depletion, etc. Hence, attention is paid to methods that either reduce electricity consumption or at least shift it away from periods of high loads (as running clothes dryers at night does). Improved sensor, computing and communication technologies make it possible to implement sophisticated schemes that were unthinkable until recently, and which promise to optimize various criteria as determined by economic models.

Similar attempts to implement fine-scale charging are apparent in other areas. This paper explores the history of the cost of charging for the use of several different transportation and communication services, and its consequences. It is not a comprehensive survey, but it does provide information that is not easily accessible in any single source. The emphasis is on the explicit costs of the charging mechanism (which are often surprisingly high, even in modern electronic toll collection (ETC) systems) as well as on the implicit costs imposed on users (such as their time or mental processing costs). We find that pricing often leads to counter-intuitive results.

There is a trade-off between the benefits of metering (charging per use) and its costs. The benefits depend upon the cost structure of the underlying transportation and communication technology, while the costs depend on the cost structure of the revenue collection technology and on the burden it imposes on users.

Individual transportation and communication technologies differ. Transportation, especially the road sector, is wrought with substantial negative externalities, most notably congestion. It is politically difficult, monetarily expensive and time-consuming to add capacity. Communication tends to be at the other end of the spectrum, with costs of increasing capacity often low and generally not objectionable to the public (with some exceptions, such as opposition to cellphone towers). Furthermore, in many cases, service providers have strong incentives to increase usage.

This paper begins with an extremely brief overview of relevant economic theory. Then, several vignettes are presented which describe the experience with pricing of different technologies (turnpikes and toll roads, the London Underground and the Internet).

A discussion and conclusion tie the various sections together, suggesting a broader and non-ideological consideration of financing network infrastructure that depends upon the underlying technology.

2. Economic theory

(a) *Fixed and variable costs*

The total cost of production generally includes a fixed cost and variable costs. The fixed cost of a road to serve one car is the same as that to serve 100 cars. The variable cost differs, particularly if congestion sets in. Combining the declining average (or per user) fixed cost component and the rising variable cost component gives a *U-shaped* curve.

The cost of collecting revenue to pay for the road has a similarly U-shaped function. A revenue collection infrastructure has both fixed and variable costs. Each technology of revenue collection will vary in the relative share of fixed and variable components. Some technologies will have a higher fixed cost and lower variable costs, others the reverse. Furthermore, some will be able to be spatially and temporally specific, while others only suitable for relatively crude identification of use in space or time.

(b) *Price discrimination*

The standard conclusion of conventional economics when applied to physical goods has been that the maximal benefit is obtained when price equals marginal cost. But that does not work well when marginal costs decrease with volume (the l.h.s. of the U-shaped cost curve). This happens frequently with information goods, and also with many communication technologies and uncongested transportation facilities. In these situations, marginal cost pricing does not recover costs and sellers have strong incentives to price discriminate, namely to charge prices that differ across different customers, depending on those customers' willingness and ability to pay. [Odlyzko \(2004\)](#) presents many examples from postal services, canals, lighthouses and other industries, where price discrimination played an important role. It was common for tolls on canals or river navigation projects to vary dramatically depending on the nature of the cargo.

A major reason for this discussion of price discrimination is that it is often the main reason behind pricing and architectural decisions taken by service providers. However, this is often carefully concealed and denied even when obvious. While price discrimination, enabled by the exploitation of monopoly power, is in many ways the Holy Grail of commerce, as it leads to maximal profit, it is a practice that arouses strong opposition (see [Odlyzko 2003b](#)). It was price discrimination by railways that brought the first serious federal regulation of commerce in the USA. Moreover, historically, differential pricing has almost always been controlled in transportation and communication. Such control is at the root of common carriage, with its prohibition on 'undue' or 'unreasonable' price discrimination.

(c) *Flat-rate preferences and stimulation of usage*

Popular resistance to price discrimination has often been reflected in legislative and regulatory constraints on the practice. But consumer resistance to such practices, as well as to fine-scale charging in general, also arises in more market-oriented ways. One of them is through a marked preference for flat rates.

It turns out that people are often willing to pay more for simple pricing. Although there have been numerous cases of this phenomenon over the ages, they tended to be regarded as irrational aberrations.

A collection of historical examples of the preference for flat rates is presented in A. Odlyzko (2000, unpublished manuscript, <http://www.dtc.umn.edu/~odlyzko/doc/history.communications0.pdf>) and Odlyzko (2001). In particular, references given therein discuss the results of careful studies carried out at AT&T in the 1970s. Three main reasons for the flat-rate preference were identified there. The first reason was the insurance effect (even if usage is low now, it might spike up as a result of some unforeseen circumstance), the second one was a systematic overestimate of usage (which made flat rates appear more economical than they were) and the third one was a harder to characterize hassle factor. Nick Szabo (see Odlyzko 2001) coined the nice term ‘mental transaction costs’ for this third factor, people’s unwillingness to be bothered with fine-grained pricing.

It should be noted that flat rates can be shown to be advantageous to sellers under some conditions even in the conventional economic model, where preferences for flat rates among customers do not exist. They are a form of bundling, and so provide a way to take advantage of uneven preferences for various pieces or transactions among users (see Fishburn *et al.* 1997).

While there is a growing literature on the preference for flat rates, there is yet another factor that is very important, but has not been studied systematically. That is the strong impetus that flat rates tend to give to usage. A collection of vignettes is presented in A. Odlyzko (2000, unpublished manuscript) and Odlyzko (2001), which show that it is very common for people, when their pricing plan changes from a metered to a flat-rate one, to increase their usage by between 50 and 200%. In many situations, this is of course undesirable. But in many others, service providers do have incentives to increase usage. This is very common in communications and occurs even in transportation, as we will see later.

3. Turnpikes and toll roads

(a) *Traditional turnpikes*

The word ‘turnpike’ comes from the technology used to enforce excludability, a pike laid across the road which would not be lifted until the traveller paid his or her toll. In England, turnpikes were developed in the eighteenth and early nineteenth centuries. They helped make the British inland transport system among the best in Europe, whereas before it had been regarded as extremely backwards. For a historical perspective, see Jackman (1916) and Pawson (1977). For recent studies that show with modern quantitative methods that turnpikes advanced the British economy, see Bogart (2005).

English turnpikes were a response to the conflict between the growing needs for better transport of a developing although still pre-industrial economy and the inadequacy of the traditional system that forced local inhabitants to devote several days per year to uncompensated labour on roads. Turnpikes were an improvement, but not a very satisfactory one. There were a variety of concerns (such as about making the *King’s highway* no longer open to all, and barring the poor) which led to concessions such as the use of ostensibly non-profit trusts to operate the turnpikes, and allowing pedestrians and certain users free

passage. Still, there were repeated riots (including the famous *Rebecca* riots in Wales in the early 1840s) and constant complaints about inefficiency and fraud. As just one example, [Adam Smith \(1776, book V, ch. I\)](#), while a supporter of improved transport, had vocal complaints about turnpikes. Smith's main concern appeared to be about inefficiency. [Jackman \(1916\)](#) discusses this inefficiency and the frequent abuses, such as inflated payrolls and outright evasion or theft of tolls. The statute books of the period were full of sanctions and prescriptions. For example, toll collection was often let for a fixed annual fee (through prescribed auction processes, which, however, were often subverted through collusion). This was supposed to avoid the fraud practised by toll collectors, but of course only moved the enforcement issue into the private sector, where the winning bidder had the same issue of making sure that proper payment was delivered. In the end, turnpikes were phased out in the second half of the nineteenth century and replaced by open public roads. For us, it is worth noting that many of the complaints about turnpikes were about the non-monetary costs to travellers (such as the delays and inconvenience of toll collection). It was also quite common for travellers to negotiate flat-rate annual contracts ([Jackman 1916](#)).

It is understandable that the old turnpikes were expensive to operate. Even disregarding fraud and inefficiency, the need to staff many toll collection stations was bound to produce high collection costs. What is amusing is that many of the modern ETC systems, which supposedly benefit from modern technologies, are also very expensive, even when one considers just the explicit costs to build and operate them.

(b) *Toll collection technology*

ETC systems are now common on many toll roads, made possible by a variety of automatic vehicle identification technologies. On toll roads, ETC increases toll lane capacity, thereby reducing toll processing time and queue lengths at toll plazas. Thus, both delays and the number of toll takers are reduced. The most advanced open-road ETC can identify vehicles at full speed, reducing vehicle delay due to toll collection to zero, from a time of 10–15 s associated with paying at a tollbooth, and eliminating the concomitant queuing when demand exceeds the available capacity of manual tollbooths. Furthermore, by reducing delays, ETC increases throughput per lane, from 350 to 400 vehicles per lane through a toll lane with manual collection up to 2200 vehicles per lane.

Prior to the advent of ETC, turnpikes in New York and Pennsylvania lost between 14 and 19% of revenue collected to collection costs using the then-current (labour-intensive) technology ([Gittings 1987](#)). This compares with 9.31% found on California's bridges, with the highest efficiency on the most heavily travelled Bay Bridge between San Francisco and Oakland. When tolls doubled on California bridges in 1998, the cost of collection relative to revenues collected was halved (aside from additional delays due to the need to give change). Hence, there is no reason to expect these percentages to remain stable as tolls vary.

Even with the adoption of ETC on conventional toll roads, when those roads remain embedded in a network of untolled free roads, many users will not adopt ETC. [Finkelstein \(2007\)](#) notes an equilibrium market saturation of approximately 60%. As a consequence, manual toll collection still has environmental consequences associated with deceleration and acceleration and remains economically less efficient than other means of collecting revenue ([Peters & Kramer 2003, 2005](#)).

Philip & Schramm (1997) have shown that ETC can reduce the cost of staffing tollbooths, money handling and roadway maintenance by 43.1, 9.6 and 14.4%, respectively. Mitretek Systems (1999) for the Oklahoma turnpike system reports that the annual cost to operate an automated lane in the Oklahoma turnpike system is only \$15 800 while the annual cost to operate an attended lane is \$176 000.

ETC has led to a modest resurgence in the use of tolling as a finance mechanism, increasing from 5.1% of total revenue in 1995 to 5.7% in 2005 (Federal Highway Administration 1995, 2005). The number of toll roads currently being planned in the USA comprises 656 km costing an estimated \$6.7 billion (Federal Highway Administration 2006). This is a significant number, but pales in contrast to the 40 000 km of the largely fuel tax-financed interstate highway already built. While some suggest a tipping point in the construction of new toll roads (Orski 2006), it is important to note that a number of roads were disturnpiked over this period, for instance tolls were removed from the Kentucky Turnpike system. Furthermore, as pointed out by Finkelstein (2007), the reduction in the salience of the electronically collected toll allows agencies to increase tolls beyond what they may have been able to do were the out-of-pocket charge felt directly.

One of the advantages of electronic tolling is the ability to impose dynamic or variable tolls, which change by time of day or level of congestion, but, as of 2006, only 29 facilities worldwide had done so (Burris 2007). This suggests that the prime motivator for moving towards electronic tolls is the reduction in transaction costs for the producer, and perhaps the ability to raise tolls with fewer complaints, rather than with improved system management.

A few places have seen congestion charging imposed primarily for traffic (or environmental) management, rather than as a way to pay for the bonds floated for the construction or to enrich investors. Notably, congestion charges have been imposed in Singapore (McCarthy & Tay 1993), London and Stockholm (Armelius & Hultkrantz 2006), which after a six-month trial in 2006 was then approved by voters in the city (though rejected by those in the suburbs) and implemented by the government. One of the key criticisms of the London scheme has been its high collection costs. The London congestion charge, which now charges the vehicle owner £8 per day to travel in the centre of London, before the recent toll increase (from £5) had enforcement and collection costs of approximately 67% of operating revenue (Hensher 2003), which compares with a less than 1% collection cost loss associated with fuel taxes (Wachs 2003). Prud'homme & Bocarejo (2005) conducted a thorough, though controversial, analysis arguing that while the scheme may be a political and technical success, it is an economic failure owing to its high toll collection costs. The authors estimate the revenue collected from tolls as 2.5 times as large as the benefits. Mackie (2005), critiquing their study, notes their findings as ironic, since the main fear about implementing congestion pricing was that it would be an economic success but political failure. It might further be noted that, since the introduction of the congestion charge, parking revenue in central London dropped (due to the reduced demand; Arup & Partners 2004). The same revenue might possibly have been raised much less expensively (and much of the same effect achieved) through the relatively efficient means of taxing parking revenue and charging for parking spaces, though that would not capture through traffic.

Another use of ETC has been the conversion of carpool (high-occupancy vehicle) lanes to high-occupancy/toll (or HOT) lanes, allowing non-carpool traffic to buy into the excess capacity of the carpool lanes. In the USA, this has occurred at five locations, one of them in Minneapolis on I-394. While again generally heralded as a success, the MnPass HOT Lanes in the Twin Cities did not recover system operating costs after more than a year in operation (Metropolitan Council 2006), though this was promised to happen shortly. It appears probable that they will never recover capital costs of investing in the additional infrastructure required to implement the system. Whether the congestion reduction benefits, or the gains from sorting drivers who have different values of time or values of reliability, outweigh those capital costs is not yet known.

The question arises as to why tolls would be preferred to taxes to pay for the construction and maintenance of roads, given their high collection costs, especially when congestion pricing of some sort is not the prime motivator. A number of reasons have been given, though Levinson (2002) poses this as a question of the ability to tax-export. Without excludability, roads may be used by both the local and through traffic. If the through traffic comprises a significant portion of the total traffic, the economic free-riding problem is quite severe. This has several consequences. The incentives for locals to pay for the road are diminished, and there is a welfare loss associated with underprovision of infrastructure. There is also overconsumption of local roads by non-residents who are not charged for their use. Underprovision and overconsumption naturally lead to poor quality. This poor quality of roads was seen prior to the imposition of turnpikes in the late seventeenth and eighteenth centuries in England.

By contrast, with tolls placed at borders, locals can offload a disproportionate share of road costs on non-residents, tax-exporting in a sense, and meeting what might be dubbed *the Monty Python test* for an ideal tax: 'taxing foreigners living abroad' (Chapman *et al.* 1989). Levinson (2001) shows that states in the USA with more non-resident travel (typically smaller states in the northeast part of the USA) are more reliant on tolls than larger states with a higher proportion of local traffic.

4. Public transport in London

Shillibeer's omnibus, started in 1829, was the first regular bus transportation in central London. It charged fares of one shilling, quite expensive for that day.

The Central Line, opening in 1900, advertised itself as 'The Twopenny Tube' in recognition of its flat fare (in contrast to the convoluted fare structures present on other railway lines and still present on surface railway systems in England today).

Paying fares for transit was historically quite common, in large part because these public transport systems were private, for-profit enterprises. The costs of collection were not insignificant. The job of the conductors was to collect and enforce fares, though their presence may have had other positive effects on passenger behaviour and aided in the maintenance of capital. On a vehicle like a bus or an electric tram, the presence of a conductor, who in 1900 was paid 4 shillings and sixpence daily, almost as much as a driver's 5 shillings per day (Harley 2002), could nearly double labour costs. Labour was a

considerable cost associated with service provision, more so on buses than on trams, which had higher capital costs but could in principle carry more passengers per driver.

On the Underground, costs associated with ticket-taking were significant. For instance, the 1926 report of London's Hammersmith and City Line had 'Ticket Collectors, Policeman, and Porters, &c' as costing £12 045 out of £29 333 of traffic expenses and compared with £85 637 of passenger receipts (Hammersmith and City Railway Joint Committee 1927, unpublished data).

As noted previously, if we believe that there are increasing marginal costs associated with use, charging fares will appropriately match demand with supply. This is the case, for example, with congested buses in peak periods. However, if we are on the l.h.s. of the U-shaped cost curve, charging fares still discourages demand, but as there are no added marginal costs with the extra passenger, this reduces total welfare. Many transit systems in the early twenty-first century are in this situation, at least some of the time. In response, many agencies sell 'season passes' under various guises (e.g. university 'U-passes', allowing students and faculty unlimited access to a transit system for a reduced fee) to encourage use. Since the pass has a one-time cost, the marginal cost for travel from the perspective of the traveller who has purchased one is zero, making the traveller more inclined to use the system. From the point of view of the operator, this increases certainty in the revenue stream, provides access to capital in advance and increases usage, while reducing the number of required transactions and thus transaction costs. From a societal point of view, this may reduce car ownership (Axhausen *et al.* 2001). Season tickets in Zurich increased bus passenger trips by 4.5%, while in other Swiss cities the increase was as large as 16% (FitzRoy & Smith 1999).

Possessing a season pass that allows unlimited rides increases usage, as the marginal cost of use once in possession of a pass drops. London has seen a major increase in public transport usage overall, with season ticket usage growing faster than ordinary payment. Between 1995 and 2005, bus ridership in London increased 40% (National Audit Office and the Audit Commission 2005). The Travelcard, allowing integrated daily (or weekly) use of all public transport in London, was introduced in 1981. London Transport claimed that the introduction of the Travelcard resulted in an increase in underground passenger-miles of 33% over the course of the decade between 1982 and 1991 (Prat 2003), though there were other conflating factors.

Many newer light railway systems (and a number of older tram systems, such as in Zurich) do not aim to check every passenger boarding to ensure payment, and instead couple an honour system with random enforcement. It is an empirical and site-specific question as to whether that trade-off is worthwhile. A system with fare machines at the entrance and exit, such as the London Underground, with fewer than 300 inspectors for 2.5 million daily passengers, attains fare evasion rates of the order of 2%. By contrast, light railway systems without enforcement estimate evasion rates that range from 1.7% in Salt Lake City to 4.7% in Denver (which has 10 inspectors and 60 000 daily trips; Leib 2007).¹

¹Lowered collection costs is one reason to rely on the honour system with enforcement and a second reason might be faster boarding times. If, as is typical with a bus, each passenger pays on boarding a light railway vehicle, the boarding time with payment is much higher. Alternative station arrangements could resolve this to some degree, which would then require more space at the station to establish a prepay and post-pay zone.

Some cities, including Seattle and Portland, have adopted fare-free zones for transit in their city centres. Other cities have made transit free city-wide, eliminating transaction and enforcement costs entirely and paying for the system as a public service out of parking fees or general revenue.

5. Telecommunications and the Internet

Many of the telecommunication technologies undergoing development and deployment (such as IP Multimedia Subsystem and Next Generation Network) appear to be motivated largely by the desire for fine-scaled charging. On the other hand, overall flat-rate plans appear to be spreading (see the comments about voice services in the papers and presentations at [ITU Workshop on The Future of Voice 2007](#), for example).

The history of telecommunications, starting with the postal system, presents a variety of examples of the importance of pricing. It also shows how conventional approaches often lead decision makers astray and often continue to do so century after century. The famous British Penny Post reform of 1840 is a very instructive example. It removed the distance sensitivity in pricing (which was motivated by price discrimination considerations and was not based on costs) and instead introduced a simple one-penny price anywhere in the UK. This led not only to a dramatic increase in usage, but also initiated a trend of a steady and relatively rapid growth in postal communication, in place of stagnant demand that had been experienced before. This is described in A. Odlyzko (2000, unpublished manuscript).

A very instructive example in considering pricing of telecommunications services is the transition in the 1880–1920 period from flat rates to metered rates for local telephone calls. This change occurred essentially everywhere around the world, with the singular exception of the USA. The motivation for such a move came from the general perception of decision makers in the phone industry, and from a variety of careful studies, which unanimously concluded that flat rates were bound to be ruinous for phone companies and unjust to users, with low-volume users subsidizing the heavy users. And indeed, with the technology of that time, where an operator was involved in setting up all calls, the marginal costs were high, so the conventional economic argument for metered rates seemed irrefutable and was not refuted. However, in the USA, this argument did not prevail, apparently owing to the competition between AT&T and the independent phone companies (see A. Odlyzko 2000, unpublished manuscript). Instead, consumer preference for flat rates led to residential pricing remaining flat rate in most of the country. In addition, contrary to the conventional economic argument, the revenues of the US phone industry as a fraction of GDP tended to be higher than those in other countries and usage, as measured in minutes per day per line, was far larger.

The Internet, which is subsuming the rest of the telecom industry, started out as a research network with no mechanism for charging individual users built in. When individual users started getting connected to it in large numbers in the early to mid-1990s, they mostly did so through commercial dial-up online services such as AOL or CompuServe, which had metered rates. Of these, AOL was the largest and most prominent, and it was a watershed for the Internet

when, in the autumn of 1996, AOL moved to flat rates. Since local calls in the USA were mostly flat rate, this meant that Internet access as a whole became flat rate in the USA. The AOL move was precipitated by the introduction of flat-rate Internet access by AT&T's WorldNet division.² AOL managers feared what flat rates would do to their network, and indeed it did get clogged, as the per-subscriber time online tripled over the following year (see the graphs in A. Odlyzko (2000, unpublished manuscript) and Odlyzko (2001)). But with time, those same managers came to appreciate the advantages of having closer contacts with customers and started to boast in their quarterly financial reports of how much usage had increased.³ Interestingly enough, AT&T WorldNet managers were not aware of the studies on the effect of flat rates and were basically looking for features attractive to their customers.

The growth of the Internet was accelerated by the switch of first AT&T and then AOL and then the rest of the industry to flat rates. This was actually a very visible and widely discussed move. Another change, also very important, took place soon afterwards, but has received practically no attention at all. It concerned wireless voice usage.

Voice telephony is unjustly neglected in policy considerations, as decision makers worldwide are preoccupied with wireline data and especially the Internet. Yet there are over twice as many wireless voice users as there are Internet users in the world today, and their spending is far more than twice as large as the spending on Internet access. (For general information on the role of both wireless and wireline voice, which still provides approximately 80% of telecom revenues, see the papers from [ITU Workshop on The Future of Voice 2007](#).) Much of the recent rapid economic growth that is credited to the Internet may instead be due to the spread of wireless voice.

In wireless (cellphone) communication, the USA is widely regarded as a laggard, with several incompatible systems and handsets several years behind world leaders. But that misses a very important point. The USA is the unchallenged world champion in wireless usage. This has been one of the most remarkable, and almost totally unknown, phenomena in communications over the last decade.

Unfortunately, unlike with wireline voice, where the International Telecommunication Union (ITU) has traditionally collected detailed usage statistics, we do not have a comprehensive database of how much time people in different countries spend on their wireless voice phones. Still, around most of the world, it appears that usage is of the order of 3–5 min d⁻¹ per subscriber. The big exception is the USA, where current (early 2007) usage seems to be approximately 25 min d⁻¹. [Table 1](#) shows the history of wireless voice usage in the USA, based on data kindly supplied by the CTIA (and available in cruder form from the graphs in [CTIA \(2006\)](#)). The USA attained its leadership position in wireless voice usage not because of any conscious policy decision by government bodies, or by the industry, but by accident. As is discussed in

²Some smaller Internet service providers had offered flat rates earlier, but they were not a serious factor in the market, and so it was only the AT&T move that forced AOL to respond.

³Interestingly, AT&T WorldNet customers did not increase their usage much when moving to flat rates, as they were interested in other activities than AOL subscribers (see [Evslin 2005](#)) and simply ended up paying extra money for the freedom from mental transaction costs that metered rates inflicted.

Table 1. US cellphone usage, minutes per day per subscriber around June of each year.

year	usage (min d ⁻¹)
1994	4.2
1996	3.9
1998	3.9
2000	7.3
2002	13.2
2004	18.1
2006	23.0

Odlyzko (2003a), the remarkable growth in usage shown in table 1 is the result of the introduction of the AT&T *Digital One-Rate* plan in the spring of 1998, which offered a block of time for a flat monthly rate, with no long distance or roaming charges. It was introduced by AT&T Wireless with very modest expectations, created in ignorance of the history of flat rates within AT&T and elsewhere in the telecom industry. But it turned out to be so popular that the rest of the industry had to follow, with the dramatic effects visible in table 1. (As of this writing, in early 2007, there are several service providers that offer truly flat-rate wireless voice service, instead of the block pricing plans that are dominant.)

Unfortunately, voice is regarded as irrelevant and hardly anyone pays attention to it. Yet it is still the dominant revenue source for the telecommunication industry and plays a key role in human interactions. One can of course ask whether encouraging more voice usage is good. But then we should also ask whether encouraging more data usage is good! And a simple response is that the increased usage visible in table 1 represents people doing what comes naturally to them when they are not encumbered by worries about per-minute billing, an evolution of phone service away from charging for each connection and towards providing a connectivity service.

From the standpoint of service providers, increased voice usage in the USA has not resulted in any disaster. The widely watched average revenues per subscriber have held up at least as well, if not better, than in Europe (see the US data in CTIA (2006)), and the industry is very healthy.⁴ This mirrors what happened with wireline voice, where the statistics in A. Odlyzko (2000, unpublished manuscript) show that unlimited local calling in the USA was associated with a generally higher fraction of GDP devoted to telephony than in other countries.

An important point about the numbers given in table 1 is that they show more than just a quick income effect. Even if one takes into account that block pricing plans did not take over right away, what happened is not that users started calling more once they were on the new plans, but rather that they gradually increased their time online. The same phenomenon occurred with the Penny Post reform in 1840 and similar reforms later in other countries. Stagnant volumes were replaced by vigorous growth.

In addition to the conventional economic arguments, sophisticated charging schemes give managers of service providers the comforting feeling that they can use them. (The technical term for this phenomenon is ‘real options’, and they

⁴Increased cellphone usage may have resulted in more traffic crashes. Eby & Vivoda (2006) review the literature, and cellphones are blamed for 1.5–5% of distraction-related crashes.

are a non-trivial concern in planning, especially in high-tech areas subject to network effects, lock-in and related issues.) This happened with the various voice prioritization options in PBXs or the many fancy features in ATM switches, which were essentially never used, but were required in procurement documents. Thus, we may very well see a proliferation of complicated control and charging mechanisms in the systems that are installed, even if they are not used in practice.

6. Discussion and conclusions

The cost of misclassifying a good with a high fixed cost and funding it on a per-use basis are several. First, there is underconsumption; by charging more than the marginal cost, society is losing welfare that could be obtained by allowing users who impose little or no burden on others from using the system. Managing this process is not easy, as undercharging may lead to overconsumption, and requires dynamic monitoring of the situation and the use of the network technology. Second, there is underproduction, when the signal to build a more comprehensive network is suppressed by overcharging. This may be especially troublesome if there exist positive network externalities.

The vignettes presented in this paper suggest that the relatively neglected transaction costs (experienced by both operators and users) may be the tail wagging the dog. Reducing these costs for consumers through bundled services (season passes, etc.) greatly increases use. Reducing the mental transaction costs through ETC (which makes the thinking about payment similar to that for a bundled good, even if every use is still charged for) reduces the salience of the charge and gives operators significant pricing power. Trying to recover costs in networked industries can be quite expensive, even with modern technologies.

The lesson for cases such as the London congestion charge is that when the aim is less to collect revenues and more to discourage usage, making the charges more intrusive is likely to be desirable. There should be no monthly passes or other easy payment methods. Requiring some positive action (involving heavy mental transaction costs) from the user, such as having to send an email or SMS message to the toll agency ahead of time, may discourage use as much as the magnitude of the toll itself.

There is no single answer that applies at all times, charging per use versus bundling depends very much on the context, the cost structure of the technology, its demand and the costs of collection for users and suppliers.

But it should be kept in mind that very often simple charging or no charging may be the ideal policy. It is not that the technology is so inexpensive that we can just give it away, that it is too cheap to meter in the words of Strauss. Rather, it is the cost of collecting charges on each transaction, both in real terms for the operator and the user and in dissuading total demand by increasing marginal costs, that makes it 'too expensive to meter'.

This research has been supported by the UK Economic and Social Research Council, and this material is based in part upon work supported by the National Science Foundation under grant no. 0236396. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Armeliu, H. & Hultkrantz, L. 2006 The politico-economic link between public transport and road pricing: an ex-ante study of the Stockholm road-pricing trial. *Transp. Policy* **13**, 162–172. (doi:10.1016/j.tranpol.2005.11.011)
- Arup, O. & Partners Ltd. 2004 Monitoring congestion charging: changes in parking usage and revenues. Report prepared for Association of London Government.
- Axhausen, K., Simma, A. & Golob, T. 2001 Pre-commitment and usage: season tickets, cars and travel. *Eur. Res. Reg. Sci.* **11**, 101–110.
- Bogart, D. 2005 Did turnpike trusts increase transportation investment in eighteenth-century England? *J. Econ. Hist.* **65**, 439–468. (doi:10.1017/S002205070500015X)
- Burris, M. 2007 Variable priced highway facilities. See <http://ceprofs.tamu.edu/mburris/pricing.htm>.
- Chapman, G., Cleese, J., Gilliam, T., Idle, E., Jones, T. & Palin, M. 1989 *The complete Monty Python's flying circus: all the words*, vol. 1. New York, NY: Pantheon Books.
- CTIA 2006 *CTIA's semi-annual wireless industry survey results*. June 2006. See http://www.ctia.org/research_statistics/index.cfm.
- Eby, D. & Vivoda, J. 2006 Driver hand-held cellular phone use in Minnesota. August 2006 report, Department of Public Safety, State of Minnesota, St Paul, MN.
- Evslin, T. 2005 Subscription pricing. February 14, 2005 blog entry. See http://blog.tomevslin.com/2005/02/subscription_pr.html.
- Federal Highway Administration 1995 *1995 highway statistics*. Table SF-21, US Federal Highway Administration, Office of Highway Investment Management.
- Federal Highway Administration 2005 *2005 highway statistics*. Table SF-21, US Federal Highway Administration, Office of Highway Investment Management.
- Federal Highway Administration 2006 *2005 toll facilities in the United States*, US Federal Highway Administration, Office of Highway Investment Management.
- Finkelstein, A. 2007 E-ZTax: tax salience and tax rate. NBER working paper.
- Fishburn, P. C., Odlyzko, A., Siders, R. C. 1997 Fixed fee versus unit pricing for information goods: competition, equilibria, and price wars. *First Monday* **2**, no. 7, July. See http://firstmonday.org/issues/issue2_7/odlyzko/.
- FitzRoy, F. R. & Smith, I. 1999 Season tickets and the demand for public transport. *Kyklos* **52**, 219–238. (doi:10.1111/j.1467-6435.1999.tb01443.x)
- Gittings, G. 1987 Some financial, economic, and social policy issues associated with toll finance. *Transport. Res. Rec.* **1102**, 20–30.
- Harley, R. 2002 *LCC electric tramways*. London, UK: Capital Transport.
- Hensher, D. 2003 Congestion charging: what Sydney can learn from London. *On line opinion: Australia's e-journal of social and political debate*. See <http://www.onlineopinion.com.au/view.asp?article=735>.
- ITU Workshop on The Future of Voice 2007 Presentations and background papers. See <http://www.itu.int/spu/voice>.
- Jackman, W. T. 1916 *The development of transportation in modern England*. Cambridge, UK: Cambridge University Press.
- Leib, J. 2007 Light rail lacks fare enforcers: why pay for a ticket to ride? *Denver Post*, 18 January 2007.
- Levinson, D. 2001 Why states toll: an empirical model of finance choice. *J. Transport. Econ. Policy* **35**, 223–238.
- Levinson, D. 2002 *Financing transportation networks*. Northampton, MA: Edward Elgar Publishing.
- Mackie, P. 2005 The London congestion charge: a tentative economic appraisal. A comment on the paper by Prud'homme and Bocarejo. *Transp. Policy* **12**, 288–290. (doi:10.1016/j.tranpol.2005.04.001)
- McCarthy, P. & Tay, R. 1993 Economic efficiency vs. traffic restraint: a note on Singapore's area license scheme. *J. Urban Econ.* **24**, 96–100. (doi:10.1006/juec.1993.1027)

- Metropolitan Council 2006 I-394 toll lane has public support. *Metropolitan Council Directions Newsletter*, July 2006.
- Mitretek Systems, Inc. 1999 *Intelligent transportation systems benefits: 1999 update*. Washington, DC: Mitretek Systems, Inc.
- National Audit Office and the Audit Commission 2005 Delivery chain analysis for bus services in England. Report by the Comptroller and Auditor General. HC 677 Session 2005–2006, 9 December 2005.
- Odlyzko, A. 2001 Internet pricing and the history of communications. *Comput. Netw.* **36**, 493–517. (doi:10.1016/S1389-1286(01)00188-8)
- Odlyzko, A. 2003a The many paradoxes of broadband. *First Monday* **8**, no. 9, September. See http://firstmonday.org/issues/issue8_9/odlyzko/.
- Odlyzko, A. 2003b Privacy, economics, and price discrimination on the Internet. In *ICEC2003: Fifth International Conference on Electronic Commerce* (ed. N. Sadeh), pp. 355–366. New York, NY: ACM.
- Odlyzko, A. 2004 The evolution of price discrimination in transportation and its implications for the Internet. *Rev. Netw. Econ.* **3**, 323–346.
- Orski, K. 2006 Highway tolling has reached the tipping point. *Innovations Briefs*, 14 February 2006.
- Pawson, E. 1977 *Transport and economy: the turnpike roads of eighteenth century Britain*. New York, NY: Academic Press.
- Peters, J. R. & Kramer, J. K. 2003 The inefficiency of toll collection as a means of taxation: evidence from the Garden State Parkway. *Transport. Q.* **57**, 17–32.
- Peters, J. R. & Kramer, J. K. 2005 Seasonal variation in the environmental effects of toll collection: evidence from the Garden State Parkway. *Public Works Manag. Policy* **9**, 287–291. (doi:10.1177/1087724X05274792)
- Philip, D. & Schramm, W. 1997 Cashless tolls mean money saved. (Reprinted for Hughes Transportation Management Systems, Canada. Dorking, UK: Traffic Technology International.)
- Prat, A. M. 2003 Demand and revenue implications of an integrated public transport policy. The case of working paper, Department of Applied Economics at Universitat Autònoma of Barcelona.
- Prud'homme, R. & Bocarejo, J. P. 2005 The London congestion charge: a tentative economic appraisal. *Transp. Policy* **12**, 279–287. (doi:10.1016/j.tranpol.2005.03.001)
- Smith, A. 1776 *An inquiry into the nature and causes of the wealth of nations*. London, UK: W. Strahan and T. Cadell.
- Strauss, L. 1954 Speech to the National Association of Science Writers, New York City, 16 September, 1954, reported in *New York Times*, 17 September, 1954.
- Wachs, M. 2003 A dozen reasons for gasoline taxes. *Public Works Manag. Policy* **7**, 235–243. (doi:10.1177/1087724X03253152)