ESRC Transport Studies Unit
University College London

THE ECONOMIC COSTS OF ROAD TRAFFIC CONGESTION

Phil Goodwin
Professor of Transport Policy

A Discussion Paper published by the Rail Freight Group

May 2004

The Economic Costs of Road Traffic Congestion

Summary

The main cause of road traffic congestion is that the volume of traffic is too close to the maximum capacity of a road or network. Congestion in the UK is worse than many, perhaps most, other European countries. More important, it is getting worse, year by year. Current official forecasts imply that congestion will be substantially worse by the end of this decade, even on the very favourable assumption that all current Government projects and policies are implemented in full, successfully, and to time. This is because road traffic is growing faster than road capacity. This is not a temporary problem: it will continue to be the case, in the absence of measures to reduce traffic, because it is infeasible to match a road programme to unrestricted trends in traffic growth.

The effect, using the current Government method of measuring congestion, and a long established method of valuing it, would be that the widely quoted figure of an annual cost of £20 billion, would increase to £30 billion by 2010.

Under current social and economic frameworks, there are no feasible policies that could reduce congestion to zero in practice, or that would be worthwhile doing in theory. But savings worth £4b-£6b a year could in principle be made by congestion charging alone, over the whole network, of which (very approximately) half might be reflected in the prices of goods, and half in savings in individuals' own time spent travelling. A good proportion of this could alternatively be secured by an appropriate package of alternative measures: priority lanes and signalling; switching to other modes including freight to rail and passenger movements to public transport, walking and cycling; 'soft' policies to encourage reduced travel by car; land-use patterns which reduce unnecessary travel; and associated measures to prevent benefits from being eroded by induced travel. The combined effects of road charging and a supportive set of complementary measures represent the best that could be reasonably achieved in the short to medium run. This could reduce congestion costs (as distinct from slowing down their increase) by 40%-50%.

These broad-brush figures, though based on long-established methods, must be treated with great caution. The 'cost of congestion', as used for these calculations, is based on relationships which in reality are not exact, stable or even meaningful. The wrong indicator has been used, comparing average real speeds with average ideal speeds. But in the real world, speeds are different every day, and so is the level of congestion. For just-in-time operation, and for much personal and business travel, variability and reliability are much more important. The really costly effect of congestion is not the slightly increased average time, but the greater than average effect in particular locations and markets, and the greatly increased unreliability.

During the near future, until road pricing is implemented, increases in road congestion can lead to some shift in the balance of attractiveness of rail freight, sufficient for a proportion of the freight market to transfer from road. This would in turn make a small but significant contribution to reducing congestion, especially in some specific important corridors. Even though rail freight is usually a small proportion of all freight, the annual economic saving in congestion cost, to road users generally, from transferring a 5-times a week, 200 mile round trip, mostly on congested motorways, from road to rail would be in the order of £40,000 to £80,000, to which should be added the commercial cost savings made by the freight operator who chooses to do so. It should be emphasised that sustaining this would require measures to prevent induced car traffic filling up the relieved road space.

An example of the impact of factoring in unreliability is given by approximate calculations made for journeys such as Glasgow to Newcastle, Cardiff to Dover, or London to Manchester. In free-flow theory these could be 3-hour journeys, but moderate congestion requires adding an hour to the average time and another hour safety margin to ensure that a tight delivery slot is not missed too often. In congestion so severe as to double the average time, the extra safety margin for unreliability could be as much as 4 hours, which is simply not feasible in many cases.

The ‘total cost of congestion’ is a large number, but it is practically meaningless and by ‘devaluing the currency’ it distracts attention from more important, achievable, objectives. It would be better not to use it as a target for policy. The two key important things to do are:

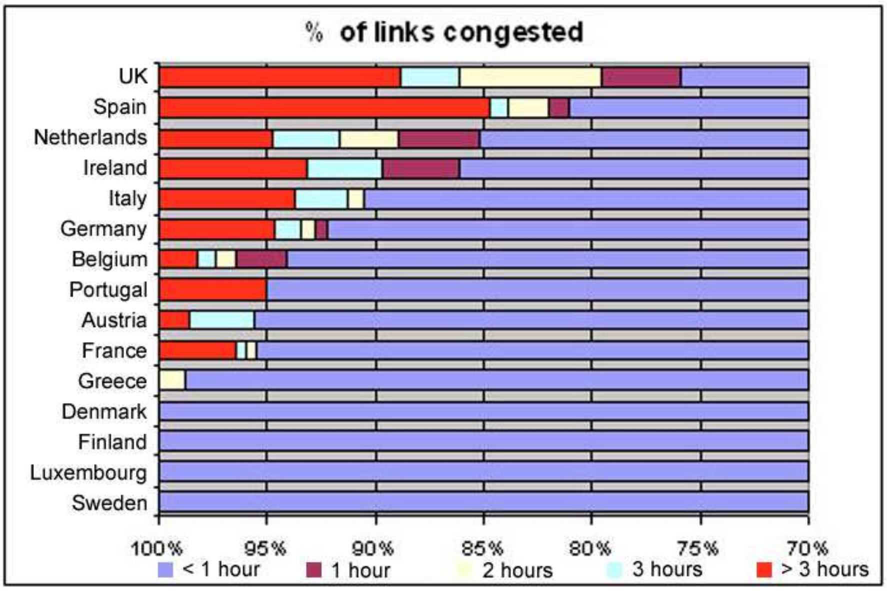
- Strategic action to reduce traffic volume to a level where conditions do not vary too much from day to day. In some circumstances this will slightly increase average speed, though not always: in some road conditions a reduction of average speed can greatly improve the smoothness of traffic flow. But in both cases, it will greatly increase reliability, this being more important than the change in average speed;

- Practical measures to provide good alternatives for freight and passenger movements which reduce the intensity of use of scarce road space in congested conditions. Even where this only applies to a minority of movements, significant effects are possible.

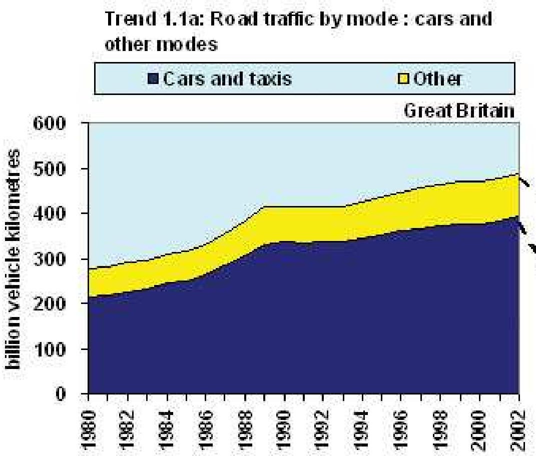
The Government plans to ‘re-launch’ the Ten Year Plan for Transport this Summer or Autumn. It is not reasonable to expect that the re-launch will include congestion charging for cars within the decade, so it will need to plan for it as soon as possible after, and a short-term coping strategy of priority measures to protect the most important classes of movement (both passenger and freight) from congestion in the period before charging is implemented.

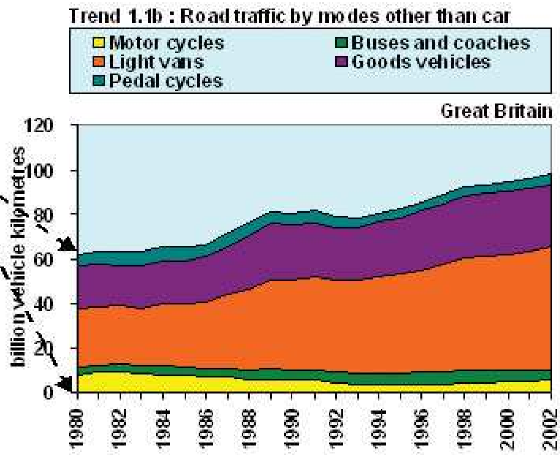
1. **The Statistical Background**[1]

Although road traffic congestion is now a feature of all developed economies, *the UK is among the worst*, as shown in a review by the Commission for Integrated Transport



By far the greatest proportion of the traffic growth has been cars, which now form 85% of all traffic on the road. Goods vehicles form 5.8% of all vehicle movement, and even allowing for their greater size this still represents only 10%-15% of the effect on traffic from the point of view of congestion. *Cars have been the biggest component of traffic growth.*



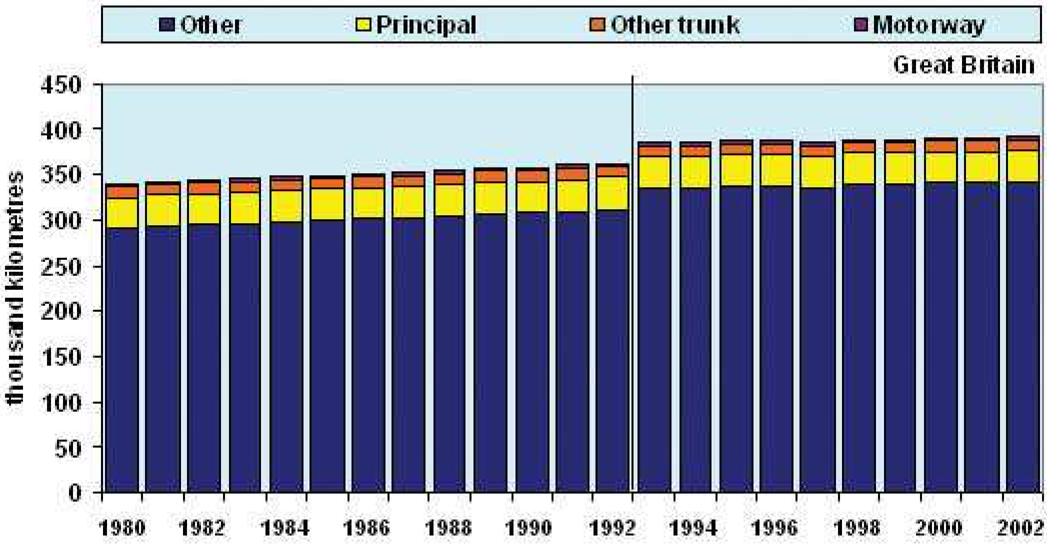


Department for Transport records show that since 1950, the amount of road traffic by vehicles in Great Britain has increased nearly tenfold – from 46.5b vehicle kilometres to 485.9 veh-km in 2002. But during this period, the total length of road in the country increased by 32%, from 297,000 km to 392,000km. Road length grew by about half a percent per year (and road capacity at perhaps twice this rate, allowing

[1] CfIT (2002) Paying for Road Use; Roads, Vehicles and Congestion, DfT London 2003; Transport Statistics Great Britain 2003, Department for Transport, London, TSO. Tables 9.7, 9.6, 1.3, 1.14. National Rail Trends 2003-2004 Quarter 2, SRA, London, December 2003 Section 3.

for motorways and trunk roads) , but traffic grew by nearly 5% a year. *The imbalance between traffic growth and road capacity has been the main influence on congestion.*

Trend 1.5 : Road length by road type



Car drivers make 28% of their trips for journeys to work, or business. The majority are a wide range of different purposes such as driving their children to school (5%), shopping (22%), and visiting friends, entertainment and holidays (24%). It is the non-work journeys that have been growing fastest. *Non-work journeys have become the biggest part of car traffic.*

Against this background, we note that the majority of freight, measured in tonne kilometres, moves by road (64%), not rail (8%), the rest going by water and pipeline. Rail freight had been in long term decline, dropping by some 60% from the 1950s to the 1980s, but has recently shown growth – the latest SRA figures show freight lifted increasing at the rate of 7% and freight moved at 1%, year on year, the turning point in freight moved being about 1995. *Rail freight has been small in comparison to road freight, and declining, but now is showing growth.*

Freight moved (billion tonne kilometres)

Great Britain 1952 to 2002-03

2. Prospects for the Economic Cost of Congestion, Using the Standard Established Methods

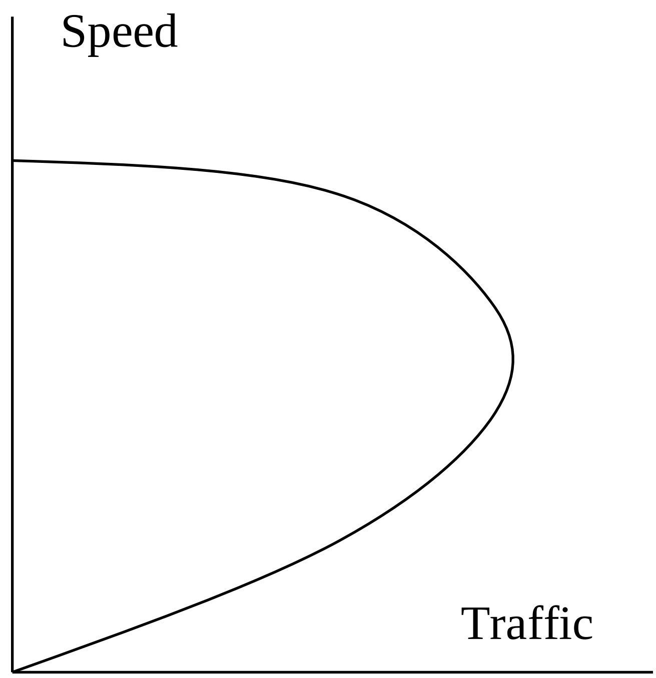
All the statistics summarised above are reasonably well-established, and uncontroversial, though the conclusions drawn from them vary widely. It is important to consider the nature of congestion, and what are its economic costs, and what to do about it.

In this section, there is an outline of the main method that has developed over the last half century for calculating the economic cost of congestion. Essentially it is based on calculating an average traffic speed, comparing that with what the speed ought to be, and multiplying the short fall by the value of time. At this stage I should say that there are some aspects of this that I do not think are helpful or logical – but those criticisms are left to the next section.

Calculating the Costs

The fundamental scientific relationship in traffic engineering for half a century has been the observation, originally called the 'speed-flow curve', that the more traffic tries to use a road or a road network, the slower it goes. Figure 1 shows the simplest useful format.

Fig 1. The Speed-Flow Curve



Thus congestion is a characteristic of all heavily used transport systems. Its general feature is that users impede each other's freedom of movement. The *general* definition of congestion therefore most usefully relates to this general property of transport systems, namely:

> *Congestion is defined as the impedance vehicles impose on each other, due to the speed-flow relationship, in conditions where the use of a transport system approaches its capacity.*

This definition indicates that the underlying *cause* of congestion does not consist of the transient and immediate triggers which drivers notice when they are in a traffic queue, such as roadworks or accidents. Nor is it based on identifying the specific scapegoats that drivers might blame in accordance with their own prejudices – women drivers, taxi-drivers, post-office vans, buses, inconsiderate parking, pedestrian crossings, police, speed limits, poorly set traffic signals, obstructions to vision. The point is that when traffic flows approach too close to capacity, *any* of the transient incidents and problems will have a disproportionate effect.

As shown on the diagram, at very low levels of traffic volume, called ‘free flow’, changes in the number of vehicles have little effect. But as traffic volume increases, even very small increases or reductions in traffic will have a disproportionately large effect on speed.

For about half a century, the method of calculating congestion costs is based on comparing the speed at free flow conditions (ie with virtually no traffic) with the speed in congested conditions (ie with the real amount of traffic), in the following steps:

Calculation of cost of congestion

(Time at ‘free-flow’ speed) *minus* **(Time at actual speed)**

multiplied by

(Volume of traffic)

equals

(Total Congestion Delays)

multiplied by

(Value of Time)

equals

(Economic Cost of Congestion).

This procedure, with some minor modifications and some more important differences in assumptions, has been used in countless studies over the years. Table 1 summarises the results of some of the best known studies calculating the ‘total cost of congestion’ defined in this way.

Table 1. Estimates of the ‘Total Cost of Congestion’

Source	Estimate	Comment
Glanville and Smeed (1958)	£125m in urban areas, £45m in rural areas, total £170m per year	No allowance for non working time to have a value
British Road Federation (1988)	£3b per year conurbations, £15b UK	History of the source of these two identical estimates is obscure – appear to make different assumptions, same answer
CBI (1989)	£5 per week per household, £15b total	
Newbery (1993)	£19.1b	Both based on similar comparison with free flow, but different treatment of marginal cost
Dodgson & Lane (1997)	£7b	
Mumford (2000)	£18b	Updating earlier figures for inflation
Smith Group (1999)	£20b	Quoted by Adam Smith Institute

Some of the differences among these estimates are due to updating for inflation, but not all. The dominant (and certainly most frequently cited) figure for the total cost of congestion is around £20b per year, equivalent to about £1000 per year per household. (The Dodgson and Lane figure has been mostly ignored by those want to quote a total cost. One the £20b had become an established number, it seems to have been thought that to use a smaller number would cause a risk of making congestion sound less important).

At face value, to a first approximation, about half of that is extra costs of business travel (freight, employees’ travel in the course of work, travelling sales representatives, meetings, etc) which directly increase the cost of production of goods and services, and therefore add to the general price level. Although this is a minority of traffic, it is usually valued at high values of time. The other half is the rather lower value of time per hour spent travelling by individuals in their own time (personal business, commuting, leisure, holidays, shopping etc), which does not directly affect prices, though it might do so indirectly by affecting wage rates and efficiency.

All this defines congestion by comparison with what would happen if the *same* volume of goods, and *same* volume of travel, were being undertaken all at the faster free flow traffic speeds.

Will it get worse?

The next step in the argument is therefore to consider the present trends. What can we say about the future prospects for congestion costs, on current assumptions, policies and economic developments?

I am not aware of a direct estimate of the *value* of future congestion costs for the UK, based on the methods used above, but there has been a substantial amount of work on the expected changes in the *amount* of congestion. The core work has been done by the Department for Transport.

The Department for Transport makes calculations of the total amount of congestion separately for each different type of area (**'a'** -cities, towns, countryside etc), types of road (**'r'** - motorways, trunk roads, lanes etc) and times of day (**'t'** - peak, off peak, inter-peak, night). This makes several thousand different cases altogether. Each of these can compare a base year – the year 2000 was used for the Ten Year Plan – and for any future year, eg 2010. From this, the percentage change in the total delays caused by congestion can be calculated, using the formula shown in figure 2. It is essentially the same principle as shown above, comparing the ideal or free-flow conditions without traffic against the actual conditions with traffic, in 2000 and 1020, and then the percentage change.

Fig 2 Department for Transport formula for changes in total congestion delays

$$\Delta\bar{C} = 1 - \sum_{art} \left[\frac{\dfrac{q_{art}^{2010}}{\sum\limits_{art} q_{art}^{2010}} \left(\dfrac{1}{\tilde{v}_{art}^{2010}} - \dfrac{1}{v_{art}^{2010}} \right)}{\dfrac{q_{art}^{2000}}{\sum\limits_{art} q_{art}^{2000}} \left(\dfrac{1}{\tilde{v}_{art}^{2000}} - \dfrac{1}{v_{art}^{2000}} \right)} \right] \times 100$$

So far, this calculation has been done twice – once in July 2000 at the launching of the Ten Year Plan, and for a second time in December 2002 for the publication of the DfT's Progress Report. (DETR 2000, DfT 2002).

Both showed (unsurprisingly) that congestion would get worse if nothing is done, but the main difference is that in 2000 the Government calculated that congestion would get *better* by 6%, from 2000, as a result of the combined effects of all its policies and projects. By contrast, in 2002 the amended forecasts suggested that even if all the plans were carried out successfully, congestion would get *worse,* by between 11% and 20%. This is shown in table 2.

Table 2 DfT Estimates in December 2002, of Increasing Volume of Congestion, 2010 compared with 2000

	Traffic Growth	**Change in Congestion**
Ten Year Plan, July 2000	**+17%**	**-6%**
Ten Year Plan Progress Report, December 2002	**+20% to +25%**	**+11% to +20%**

These figures are expressed as percentage changes in estimated time losses, not value as such. But an approximate conversion to value can be made by considering that (again on the conventional assumptions[2]) the value of time is thought to increase at or close to proportional to GDP, so we might grow the value of time by, say up to 2.5% per year in real terms, in a decade it would be 20% to 30% higher than in 2000.

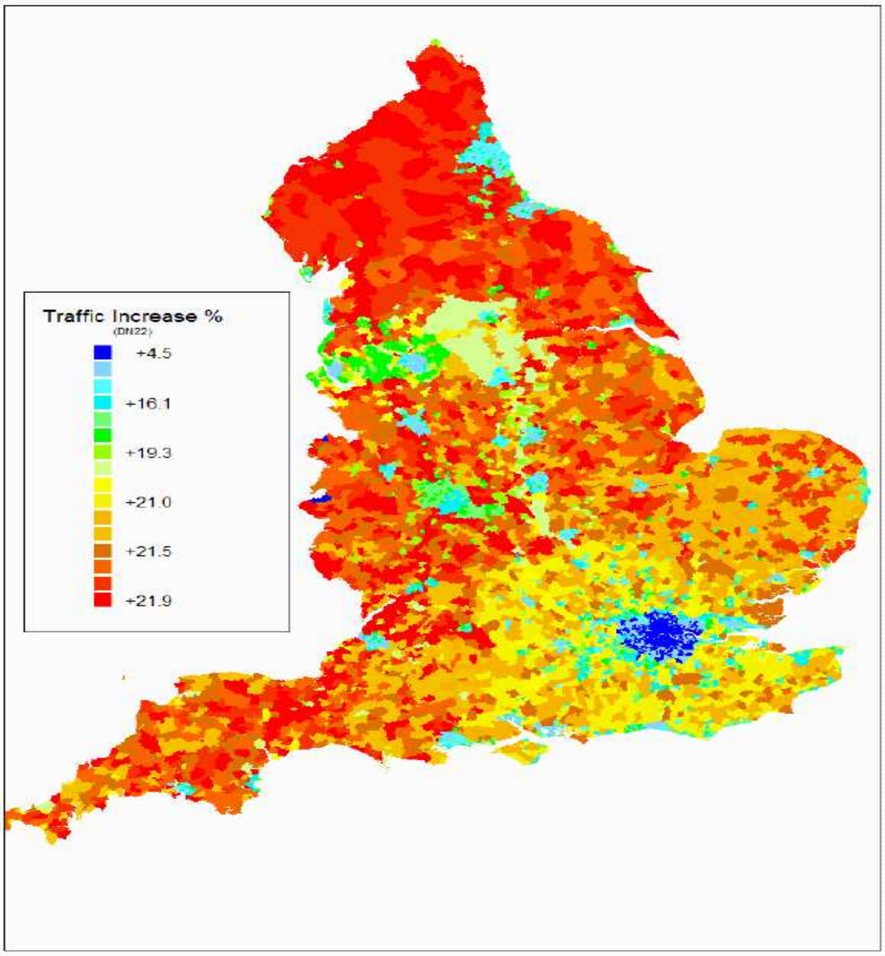
This picture is rather robust, and in broad terms has been found by many other studies, carried out by the RAC Foundation, the Highways Agency, international studies by the EC and the European Conference of Ministers of Transport, the Commission for Integrated Transport, the present author (Dargay and Goodwin, 1999), and Graham and Glaister (2003) of Imperial College, whose calculations are shown on the maps overleaf, reproduced by their kind permission. The first map shows the parts of the national road network which would have the largest increases in traffic – nearly everywhere except where congestion is currently at its highest, and there is little scope for further growth. The second shows the corresponding reductions in speed, where even small growth in traffic has a particularly heavy effect on the currently most congested areas.

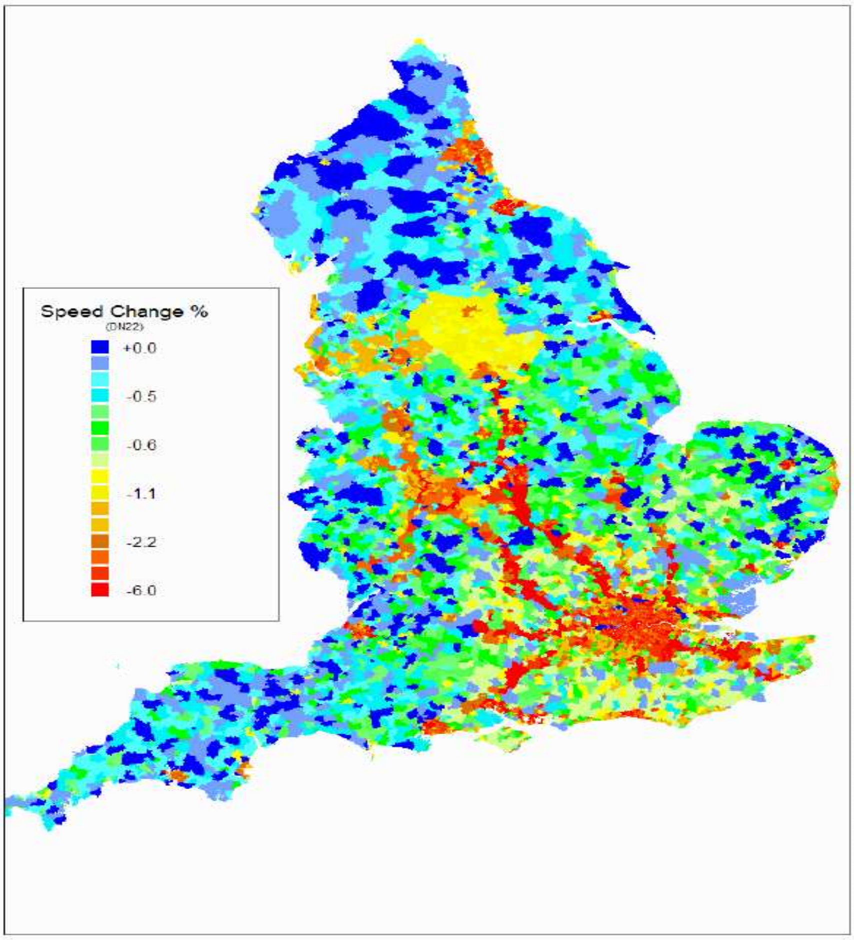
So, to a first approximation, the established methods and calculations suggest the following calculation:

Cost of Congestion in 2010 = (£20b per year) plus (11% to 20%) for congestion plus (20% to 30%) for value of time, ie £26b to £31b, or in round terms

> ***On current assumptions, the cost of road traffic congestion will be around £30b per year by 2010 – and still increasing.***

[2] There is evidence that the value of time, at least for non-business travel, grows less than proportionally to income, which would reduce this figure, and would also mean that travel time spent for goods transport would become a bigger element in cost-benefit assessments in future, compared with personal travel.


Traffic Increase %
(DN22)
+4.5
+16.1
+19.3
+21.0
+21.5
+21.9


Speed Change %
(DN22)
+0.0
-0.5
-0.6
-1.1
-2.2
-6.0

3. How Realistic are These Estimates?

At this stage, the argument seems to establish that congestion is a big problem, with great costs, and it is getting worse. To avoid misunderstanding in what follows, my judgement is that these general conclusions are true, and robust.

However, there is no specific conclusion that follows from that, in terms of policies to be favoured or rejected. The statement usually only serves as little more than a preamble to dramatise any policy being advocated by any party, and indeed, the fact that some of the assumptions and methods of estimating the total cost of congestion are challengeable, does not much matter to the solutions which follow. Indeed, the large numbers may have done a disservice to the argument, devaluing the currency of debate because by contrast the potential benefits of any real policy are likely to seem small by comparison.

To become useful, rather than simply dramatic, I suggest that two very important changes need to be made to the methodology. The first is to abandon the idea of the total cost of congestion and replace it by the marginal costs and benefits of specific interventions. This gives smaller numbers, but they are usable. The second is to replace the emphasis on average conditions by an emphasis on the variations which make journey times unpredictable. This would largely replace the 'average value of time' by the 'value of reliability'. (Note that this is not the same as the suggestion simply to increase the value of time by a factor, typically up to 50%, for travelling in variable conditions).

Critique of the 'total cost of congestion'

In recent years there have been some low-key but important doubts about whether the measure of congestion used has any real meaning. This concern is usually expressed in the form 'the measure of congestion should be meaningful to drivers' – the Secretary of State for Transport, the House of Commons Transport Committee, the Commission for Integrated Transport, and the CBI have all used this phrase, in almost exactly the same words, though saying little about what it means.

The problem is not about the complication of the equation shown above. It is the underlying concept of the 'total cost of congestion', which on examination becomes very odd. It is based on comparing a real-world condition with a 'target' of what the world would look like with all its present traffic, but free-flow speeds[3].

This is impossible, ever, to observe, because it is a construct which logically cannot exist outside a computer model: consider, for example, the case where somehow enough road capacity were produced to allow all the present traffic to travel at free-flow speeds, even after allowing for induced traffic. It is quite certain that the amount of traffic would increase. So the specific combination of the present quantity of traffic, plus free flow conditions, could never actually both exist at the same time. It is

[3] Strictly, the 'target' is not the pure free flow speed, but is modified by the speed limit, since in cost-benefit analyses a value is not accorded, except by mistake, to the benefits of breaking the law. An analogy might be the distress and loss of standard of life suffered by burglars if they are prevented from burgling. However real, we do not count this in a cost-benefit analysis suggesting that, overall, the economy might be better off if we encouraged them to continue.

a purely notional idea, not a conceivable description of a world we might choose to provide for.

But even if it could, this measure of congestion used has strange, and in some cases perverse, implications. Four examples will be given.

> *Perverse implication Case 1:* Consider the case if the target speed increases. We start with peak-period traffic travelling at 20k/hr on a 30k/hr local road. Then we widen the road and solve some bottlenecks, increasing the average peak period traffic speed to 25k/hr. At the same time we then redesignate the target or free-flow speed to 60 k/hr.
>
> According to the formula, congestion costs are now greater, even though in fact every vehicle is travelling faster. This is because the shortfall from 25 to 60 k/hr is ‘worse’ than the former shortfall of 20 to 30 k/hr.
>
> *Perverse implication Case 2:* Consider what happens if the volume of traffic grows. In this case, the total cost of congestion increases in proportion, even if there is no change in speed at all - or even if speeds increase, but by less than the volume of traffic. So the calculation would say, implausibly, that a growing volume of traffic, using a continually improved road system, at continually increasing speeds, could still be suffering an increased total cost of congestion. What Government would sensibly choose to describe such an outcome as economic costs getting worse?
>
> *Perverse implication Case 3.* Consider the case if we deliberately reduce the accepted ‘target’ speed (as is widely done in speed restrictions), or if we alter road design in such a way that the free-flow speed of traffic falls (as is widely done in traffic calming). Then the average speed of traffic may fall, but the calculated total cost of congestion to the economy would be less, not more[4].
>
> *Perverse implication Case 4.* At low or average levels of congestion, very small speed differences are magnified by the arithmetic into alarming changes in congestion: the ‘6% improvement in congestion’ envisaged in the Ten Year Plan translated into a few seconds, and on some road classes fractions of seconds, per mile. For reasons discussed below, the random day to day variation in speed can be much greater than this, so differences in congestion can be proposed which are virtually impossible to detect.

In summary, statements of the form ‘congestion costs the economy £20 billion a year’, updated from time to time by inflation, are good for headlines, in dramatising a large problem. But the implied annual dividend of £1000 waiting to be distributed to each family is a fiction. It is calculated by comparing the time spent in traffic now, with the reduced time that would apply if the same volume of traffic was all travelling at free flow speed, and then giving all these notional time savings the same cash value that we currently apply to the odd minutes saved by transport improvements.

[4] This may not be as silly as it sounds. Sometimes reduction in traffic speeds can improve congestion, by smoothing out the flow. But that is a different effect, considered below.

But this could *never* exist in the real world – not for reasons of practical difficulty, but because it is internally inconsistent. If all traffic travelled at free flow speed, we can be quite certain that there would be more of it, at least part of the time saved would be spent on further travel, and further changes would be triggered whose value is an unexplored quantity. It is an apparently precise answer to a phantom question. It really does not matter whether 'the answer' is £7 billion or £23 billion.

What matters in practical terms is the *change* to the cost of congestion brought about by a specific feasible project or act of policy. These numbers will generally be smaller than the famous £20 billions, but real. As economists would say, we need to change our thinking from *total* costs, to *marginal* costs.

Moving from 'Total Costs of Congestion' to 'Marginal Costs of Congestion'

Interspersed among the calculations of total costs, is a separate literature concerned with marginal costs, typically in the form of 'the effect on congestion of one extra vehicle' or 'the effect on congestion of a particular realistic policy'.

Recent work at the Institute for Transport Studies, Leeds University (Sansom et al, 2001) has reviewed evidence on such marginal costs, expressed as the effect of an extra vehicle kilometre.

Rounding their 'low' estimates (in this study there was very little difference between their low and high figures), they give figures as shown in table 3.

Table 3. Marginal External Costs of Congestion 1998 London and Conurbations

Category	Percent of national vehicle km	Costs Pence/veh km
Central peak	1%	86
Central off-peak	3%	47
Non-central peak	4%	23
Non-central off peak	8%	11
National total	100%	10

They also disaggregate, for example giving an estimate for 'inner London motorway' on a Saturday of 15 pence per pcu km.

The Leeds authors re-estimate and update earlier figures suggested by Newbery, though with some caveats. Their figures are higher than Newbery's – for example, their 11 pence/km compares with their updated Newbery figure of 5 pence, and their figure for peak period urban central areas is also greater, albeit with a somewhat different definition of size of urban area which reduces comparability. The Leeds authors comment

> 'As the figures in Newbery (1990) are based on the same methodology as that in this study, the main reasons for the larger marginal congestion costs in this study is the growth in traffic over time. The Newbery (1990) figures were based on 1985 traffic data and in the period to 1998 traffic growth and changes in speeds have been substantial'

They comment that the orders of magnitude are similar to advice given by the DETR of values to be used for assessing the decongestion benefits of 'major rail-based urban public transport'.

Table 4. ITS 'Low' Estimate of Marginal Congestion Costs for London

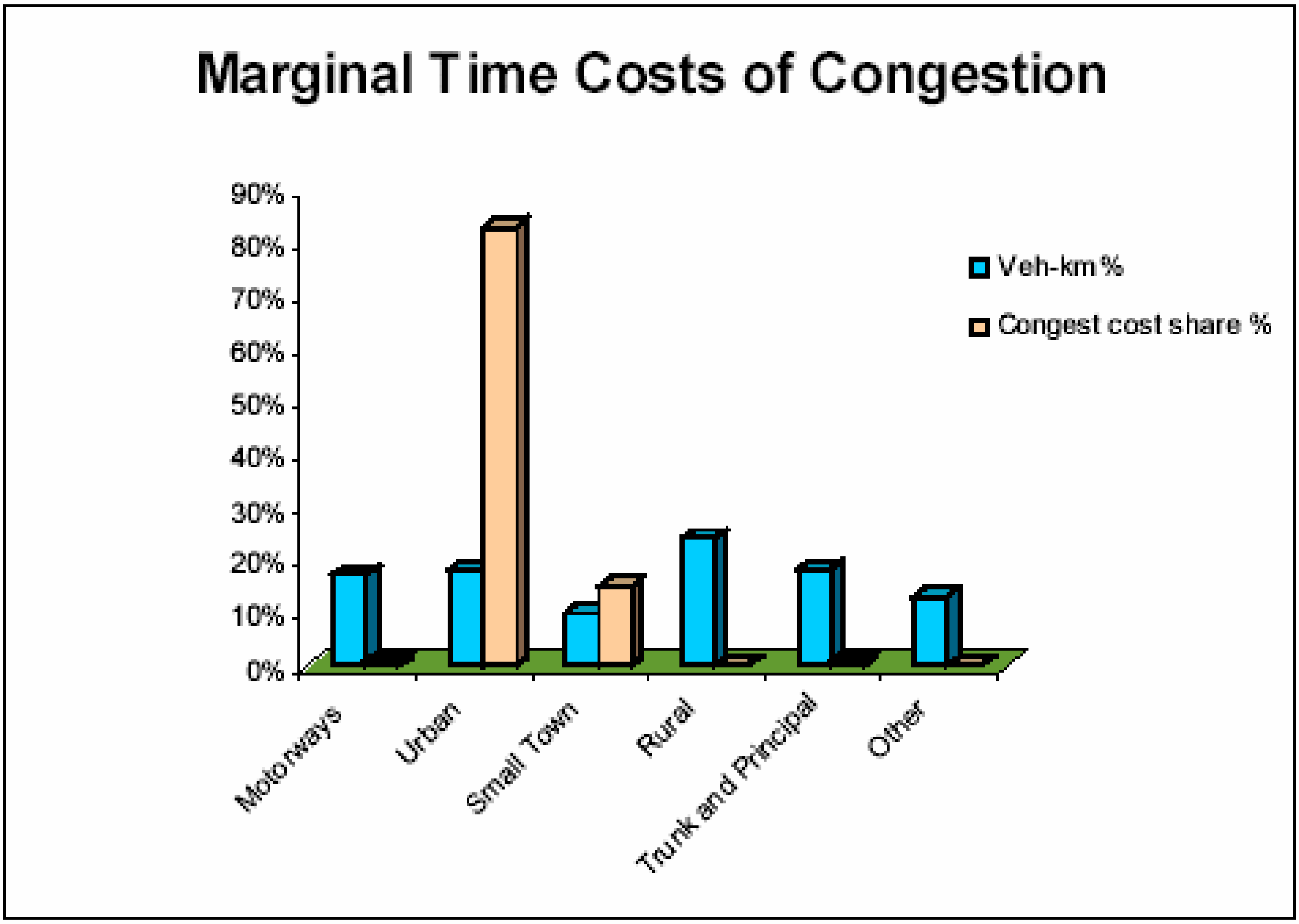
Area and Road type	Congestion cost Pence per car km
Central London	
Motorway	54
Trunk and principal	71
Other	188
Inner London	
Motorway	20
Trunk and principal	54
Other	94
Outer London	
Motorway	31
Trunk and principal	28
Other	40

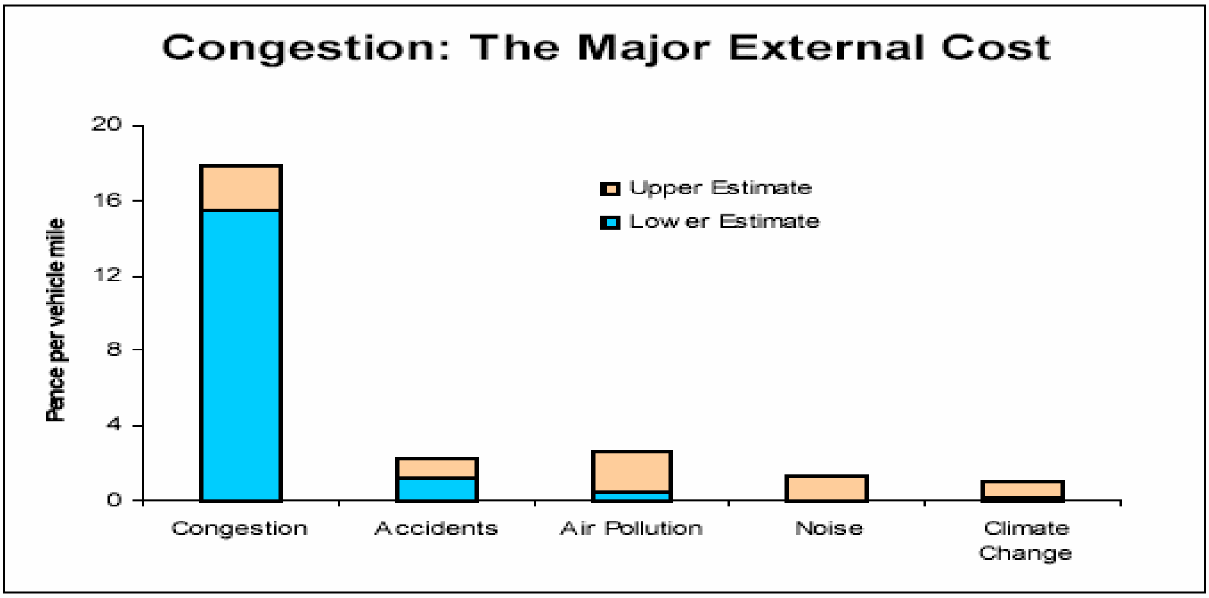
Their congestion costs are overwhelmingly greater than the sum total of their other external costs (accidents, air pollution, noise, climate change) and infrastructure costs: all these other factors barely add up to more than a penny or two per vehicle kilometre. (Some of these cost estimates are of course subject to considerable discussion, in particular in suggesting that environmental costs are higher[5], but that is outside the scope of this paper). The marginal revenue in fuel duty (and VAT on it) is estimated at 4.5 pence per vehicle kilometre on average, so broadly we can say that fuel tax provides revenue in the order of 3% to 25% of the cost of the extra congestion each vehicle causes.

The broad picture of the Leeds results are shown in figures 3 and 4: their proposition is that congestion is far the greatest part of all external costs – much greater than the costs of pollution – and the largest part of it is in towns.

---

[5] For example, SACTRA (1999) cite four studies, from Pearce 1993, Mauch and Rothengatter (1995) Royal Commission on Environmental Pollution 1994, Maddison et al 1996, giving estimates for these non-congestion elements of external cost varying from £8 billion per year to £40 billion per year, which is broadly the same order of magnitude as the total congestion costs), which, notes SACTRA 'vary from c£7 billion per annum (NERA, 1997) to c. £19 billion per annum (Newbery, 1995). These studies therefore accord broadly the same economic importance to accidents, noise, air pollution and climate change, taken together, as to congestion. This does not come out from the Leeds analyses, but intuitively seems to correspond better with public perceptions of the importance of such issues. If environmental costs were increased it would generally reinforce, rather than undermine, the conclusions of this paper





Source: "Surface Transport Costs and Charges: Great Britain 1998", Institute for Transport Studies, University of Leeds.

It follows from the marginal approach, that the policies and projects which will have the greatest beneficial effect on congestion will be specific measures aimed at specific locations, especially the most congested ones. These measures have included traffic management procedures including priority lanes and signalling; switching to other modes including freight to rail and passenger movements to public transport, walking and cycling; ‘soft’ policies to encourage reduced travel by car, and land-use patterns which reduce unnecessary travel; and associated measures to prevent benefits from being eroded by induced travel. There has also been a recurrent interest in the effect on congestion of road pricing (which recently has come to be called ‘congestion charging’), as shown by table 5.

Table 5 Economic Value of Congestion Reduction from Road Pricing

Source	Effect
London	
MoT (1964)	£100m-£150m saving
CIT (1990, 1992)	£250m saving
House of Commons Transport Committee 1994	£100m-£300m benefit
ROCOL (2000)	£95-£160m benefit
CfIT (2001)	£191m specifically for freight
TfL (2003) interim	£180m (1^{st} year costs £130m)
England[6]	
CfIT (2003) as cited in	£2.3b
Glaister & Graham (2003)	up to £4b

The notable point about this table is that the benefits achieved are typically in the order of £200m a year, for central London, and up to about £4 billion for England, giving say £5b for the UK, or £6b on less conservative technical assumptions – and this from direct intervention, with the most powerful economic instrument yet devised, with prices actually related directly to the congestion cost itself.

At first sight that seems small by comparison with the ‘total’ cost of £20 billion a year, but the figures are not comparable. The point is, we can have the £4b-£6b if we want, but the £20b is forever outside our reach. Another way of saying this is that it would not be economically efficient to try to operate at zero congestion.

The marginal approach (unlike the ‘total cost of congestion’) then enables specific useful calculations to be made of the contribution that might be made to road

[6] CfIT daytime weekdays only, Glaister and Graham at any time. (They do not cite this actual figure – I have calculated it from their results). CfIT calculations, made by Dodgson, give £1.8b value for changes in average travel time, described as a 44% reduction in congestion, which is distant from a £20b total but not far from Dodgson & Ware’s earlier £7b. Both figures depend heavily on DfT speed flow relationships, which are not sensitive enough at very high levels of congestion, and the CfIT figure relies on elasticities which have since been revised upwards and would give a higher benefit. My guesstimate is that the two studies, re-done with more sensitive recent values, would both give a figures of up to £6b for England and £7-£8b for UK. It will be interesting to see if current DfT work confirms this.

congestion by a shift of freight from road to rail. So far, the most detailed estimates made – albeit still somewhat controversial, are those made by the Strategic Rail Authority (2003), revising earlier figures they had developed over some years, as shown in table 6.

Table 2: SRA proposal for values of external benefits of rail and road freight (pence per lorry mile)

	Motorway			London & Conurbation		Rural & Urban		Weighted Average
	High Congestion	Medium Congestion	Low Congestion	Trunk & Principle	Other	Trunk & Principle	Other	
Accidents	1.5	1.5	1.5	3.8	3.1	3.8	3.1	2.9
Noise	4.0	4.0	4.0	11.0	9.0	2.0	4.0	3.8
Pollution	5.7	5.7	5.7	18.8	22.8	3.9	4.8	6.3
Climate Change	2.7	2.7p	2.7	2.6	2.5	2.4	2.0	2.5
Infrastructure Costs	5.7	5.7	5.7	9.1	28.7	11.2	35.3	12.5
Road Congestion	79.0	37.0	6.3	121.9	135.5	45.8	10.6	43.9
Unquantified	8.0	8.0	16.0	8.0	9.0	21.5	22.0	16.9
Taxation	-29.0	-29.0	-29.0	-29.0	-28.0	-29.0	-28.0	-28.9
Rail Costs	-8.8	-8.8	-8.8	-8.8	-8.8	-8.8	-8.8	-8.8
TOTAL	**68.8**	**26.8**	**4.1**	**137.5**	**173.8**	**52.8**	**45.0**	**51.1**

What the figures suggest is that – for congestion[7] alone – a single lorry mile removed from the road network by transfer to rail, would give economic benefits of up to £1.35, depending on the level of congestion, and under current conditions the overall average would be 44 pence benefit. Converting this to an annual equivalent, a 5 times a week return journey of 200 miles, transferred from road to rail, would produce a reduction in the economic costs of congestion of over £40,000 per year in average road conditions, and nearly double this, around £80,000 per year, if the transferred journey had been on congested motorways which is likely to be the case for a large proportion of traffic realistically in scope.

As traffic grows, and congestion increases, the benefit of doing so will increase over these figures. (Also, if the environmental values are increased, as suggested in the footnote on p15, the advantage per lorry mile transferred would be greater. Warren, Potter & Parkhurst (2004) estimate that if all lorries were charged the full social marginal cost of their journeys, there would be in the order of £8b revenue to the Treasury, approximately of the same order as the shortfall expected as better fuel economy and engine design for road vehicles leads to reduced tax revenue from fuel over the next two decades).

Thus the marginal approach at first sight seems to lead to undramatic figures compared with a total approach – pence per mile instead of billions of pounds per year. It does not lend itself so well to media headlines. But it is more useful, more realistic, more achievable.

[7] Note that this does not include the user benefits, such as reduced costs, which would be enjoyed by the operator who chose to make this switch for reasons of advantage in private costs. Also note that the figures for environmental costs are lower than some other estimates, but that is outside the context of this paper.

A Further Caveat

The wider benefits depend on *sustaining* the reduction in congestion brought about by such a transfer, against the tendency for traffic growth – itself reinforced by any improvement in traffic speed – to fill up the road space liberated by such policies. The problem here is that the higher the level of congestion, the greater the amount of 'latent' traffic is likely to be, and therefore the shorter the period of relief from congestion provided by a voluntary shift in choice from some sections of the market. In this sense, traffic reduction by voluntary choice, unsupported by other measures to prevent induced traffic, suffers from exactly equivalent disadvantages as the traditional solution of attempting to provide enough road capacity to keep up with demand.

However, the benefits to the transferred traffic itself do not depend on this caveat. If the relative costs of road and rail freight, for example, are such as to encourage an operator to shift from road to rail for reasons of commercial advantage, those private savings are not eroded by induced traffic filling up the space vacated on the road network. Indeed, they are likely to grow if the growth of road congestion is not prevented.

The less is done to protect such benefits on the roads, the more important does it become to cushion a proportion of goods movement from road congestion, whether by reliable rail services or other instruments of protecting this important category of traffic.

Current plans for road charging

This conclusion is further reinforced by consideration of current plans for changes in the charging regime for roads. Following the experience of congestion charging in central London, the Government is now reviewing the possibility of a nationwide congestion charging scheme, with an implied time scale of starting some time after 2011, perhaps 2015. There are also plans to implement distance-based charging for lorries, to replace current taxation systems, some time after 2006. This will not initially be variable according to congestion levels but could easily become so.

There is therefore likely to be an intervening period, in the later years of this decade and first years of the next, during which there will be a closer connection between lorry use and the costs charged for it, but not for cars. This is likely to have three effects:

- foreign lorries will be charged, and the scheme is revenue neutral for UK fleet, therefore the foreign fleet costs will increase, and the *average* price for road freight overall will increase;

- charges will be more closely based on use: this is *not* revenue neutral for any specific company so is likely to trigger more rational logistics, perhaps with some reduction in lorry road mileage travelled;
- Somehow the charging technology will need to be paid for. It is not known how much this is since the technology is not yet chosen, but McKinnon (2004) suggests that it is likely to be substantial.

Taking these together, it is likely that UK hauliers will be in a relatively better competitive position versus foreign hauliers (though this might be offset by similar changes in charging regimes abroad). Rail freight is likely to be better placed versus road freight (depending on the degree to which rail strategy in UK allocates priority to competing for the freight market). And goods transport overall is likely to be *worse* placed versus passenger movement, since congestion will increase for all, but freight will have cost increases as well.

This raises the question of the 'gap years' in transport policy for freight: what will happen between the period when distance based charging is implemented for lorries, and congestion charging is implemented for cars? The problem here is that, from a company point of view, it is worth paying more if it will reduce congestion, but it will only reduce congestion if it is applied to cars as well as lorries

The Government plans to 're-launch' the Ten Year Plan for Transport this Summer or Autumn. It is not reasonable to expect that the re-launch will include congestion charging for cars within the decade, so it will need to plan for it as soon as possible after, and define priority measures for goods transport (road & rail) *before* charging

Moving from 'Average Speed' to 'Reliability'

All the previous discussion has suffered from a convenient simplification that is almost universal in transport forecasting, but seriously reduces its usefulness. It deals with the average speeds that apply to *average* more and less congested conditions.

Although the speed flow relationship is basically a very simple proposition, and squares with everyday experience, the version shown in figure 1 is not very realistic. (And the simplified straight-line version used in many traffic models is even less realistic). In practice, the world is not so tidy, and data show that speeds can vary very widely even at the same level of flow on the same category of road. The following complications are usually not taken into account.

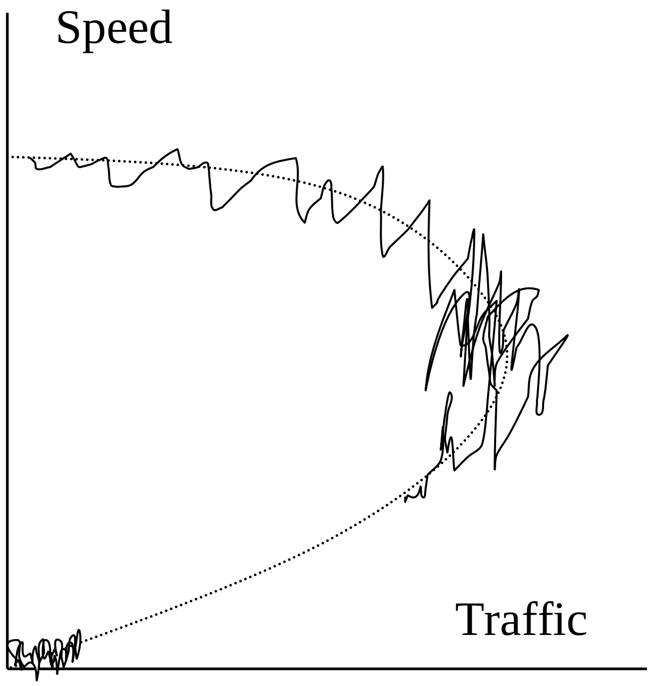
- At low levels of traffic, a small additional number of vehicles does not make much difference. But at high levels of traffic, as the maximum capacity is approached, a small change in the number of vehicles has a quite disproportionate large effect on speed. This change in the slope of the curve is particularly evident at high of traffic volumes.
- Conditions very close to the maximum capacity are not just steep: more important, they are very unstable. When this happens, the whole system becomes rather unpredictable, and speeds can vary quite widely and quickly, often without very apparent causes.

- If too much traffic tries to use the system, it fails, and flow can cease – ie, for a period, there is zero speed. This is often called 'gridlock'[8], which is not strictly accurate.

It will be seen from this that the most sensitive and difficult issue concerns the unstable, variable behaviour of traffic in the region where the road is operating close to its maximum capacity. This, unfortunately, is becoming all-too common, and it is not represented well by the averaging process of most traffic models. (On average, any variable will by definition be operating at less than its maximum).

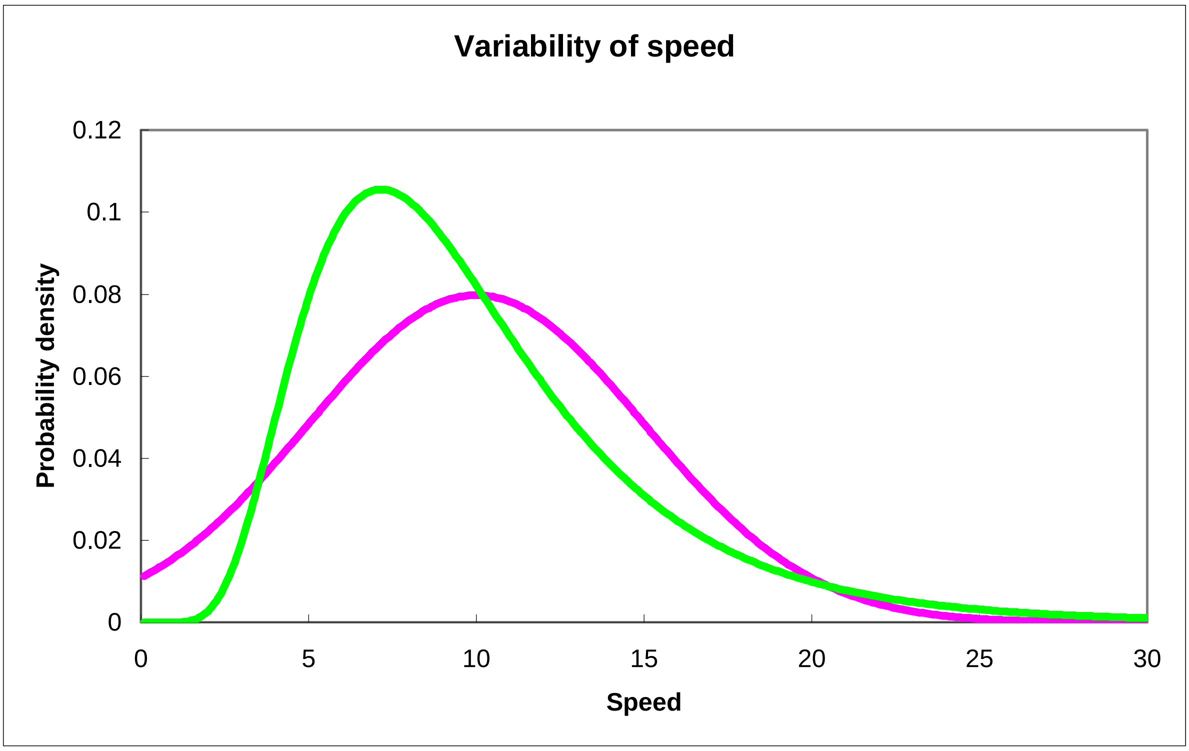
Putting these together, figure 5 is an impressionistic version of the speed flow curve that might be more realistic for real road conditions, and figure 6, courtesy of my colleague Professor Heydecker, shows examples of families of distributions of speeds, of which we would expect a somewhat different shape and numbers for each type of road for each type of weather, for each type of traffic composition, and differently for more and less congested conditions.

Fig 5: More Realistic Speed-flow Curve showing Instability and Unreliability



[8] Strictly, 'gridlock' relates to a road network organised on a grid pattern, such as some US cities, when traffic cannot move because each direction permanently blocks the other. In historic cities this rarely happens because there are always some escape routes, side streets, or illegal movements that allow a safety margin of comfort, a form of 'redundancy' or inefficiency that is essential to prevent minor problems getting out of hand. Attempting to operate permanently very close to maximum capacity, in a system subject to random variation, is asking for trouble.

Figure 6 For any average speed there is probability of a variable range of speeds



Although the mathematics then become very difficult – and the data requirements much more extensive – the general picture corresponds much more with everyday experience. The key problem is that as congestion increases, journey times are not only longer, they are also more unpredictable. In many cases, the unpredictability is a very much more important economic cost than the average speed, because it requires spending more time than ought to be necessary, as a margin of error.

Until recently, research on this aspect has been of fringe technical interest, and even when carried out it has been completely ignored in forecasting and appraisal.

The classic work on this subject is by Smeed and Jeffcote (1971). They reported on the variability of a car journey repeated 253 times between Bray and Central London, and found day-to-day variability in the overall average journey speed such that the standard deviation was between 20% and 33% of the mean. Mogridge and Fry (1984) repeated the experiment for 172 journeys between Clapham and Central London, finding a standard deviation of 15% to 20% of the mean. Mohammadi (1997) analysed data from three much larger experiments, totalling about 1300 journeys. Standard deviation for different subsets varied from 2% to 51% of the mean, the overall figures being 16%, 16% and 20% for the three experiments (and subject to influence from a variety of factors such as incidents, weather, etc). Each of these authors cites further references, with broadly similar results.

Thus, it is the experience of drivers using the same route at around the same time on successive days that a degree of day-to-day variation is part of their normal lives. For a hundred journeys to work, broadly between 10 and 20 will be travelled with an overall door-to-door time more than 20% faster than the average, and a similar proportion will travel more than 20% slower than the average. But even this variation is itself not stable from day to day – when the weather is bad, or when traffic flows are higher than usual for some random effect, then the unreliability will be higher than on 'normal' days. So we have various levels of experience – the 'average' day which is a statistical calculation, the 'usual' random variation even when conditions are rather similar, and the 'special' variation in unusual conditions.

Two consequences follow.

(a) Measures which change journey speed in the order of 5%, say - *though entirely real* - may not always be revealed statistically by comparison of one day 'before' and one day 'after' traffic observations of speeds or flows. Such considerations led, for example, to problems in detecting statistically significant changes in traffic flows resulting from the implementation of a number of traffic calming measures in Reading, because the natural variability of traffic counts was as much as 40%. (Ward, 1997). This is a real problem for national monitoring, since the very small changes in average traffic speed implied from year to year in the Ten Year Plan are very much smaller than the day to day variation that happens anyway, so difficult, or perhaps impossible, to detect with confidence.

(b) Drivers accustomed to variation in their own travel conditions are unlikely to all respond immediately to changes in speed, since such changes are not immediately obvious. Their ability to detect changes and respond to them will be influenced by how long it takes them to build up a 'true' (to themselves) picture of the average conditions, or the frequency of unacceptable journeys, depending on the criteria they use.

The research necessary to complete this picture has not yet been done. But we can make a reasonably good qualitative estimate of what it will show.

An example can be given with the following assumptions. We consider journeys of 210 miles (because it makes the arithmetic easier) – say, roughly, from Glasgow to Newcastle, Edinburgh to Liverpool, London to Manchester, Cardiff to Dover, or Newcastle to Birmingham. The target is to arrive in time for a just-in-time delivery slot, which allows a flexibility of up to ten minutes. In this example we assume that there is allowance for missing the slot on not more than one day in ten: if the allowance is stricter than this (eg only one day in a hundred), then a greater amount of spare time needs to be built in.

Then we consider three cases. First, there is no congestion and a vehicle can keep to the maximum speed limit in motorway conditions for the whole journey, with virtually complete reliability, so the average journey time is 3 hours (at 70 mph) with little or no variation. Second, there is congestion which increases the average journey time to 4 hours (52.4 mph average), with a standard deviation of journey time equal to 20% of the mean. Third there is heavy congestion raising the average journey time to 4 hours (35 mph average), with a standard deviation equal to 40% of the mean.

Table 7 shows the results, rounding off to avoid giving a spurious idea of accuracy.

Table 7 Sample calculations of effect of unreliability as congestion grows

Conditions	Average speed	Average travel time	Allowance for unreliability	Total time
No congestion or hold-ups of any kind	70 mph	3 hours	A few minutes	Close to 3 hours
Congestion and 'normal' unreliability	52.5 mph	4 hours	1 hour	5 hours
Heavy congestion, very variable and unreliable	35 mph	6 hours	4 hours	10 hours

In this case, we see that at heavy congestion, while the increase in average journey time is substantial, it is the extra allowance for unreliability that is by far the bigger effect. In these conditions, it simply becomes impractical to run according to just-in-time principles, and the entire logistical chain would have to be re-thought.

In principle, this argument leads to the following conclusion: if the Government's forecast of an increase in congestion from 2000 to 2010 is correct, when measured in terms of average speed as is currently done, then this *also* implies a greater than proportional increase in unreliability. The figures are unlikely to be as great as in the hypothetical table above, not least because if they were, individuals and companies would be forced to change their patterns of life and logistical chains in such a way as to protect themselves against the worst effects. But the effect would be larger than the direct congestion cost, and increasing.

Current work

The Department for Transport is undertaking a programme of research into variability of travel speeds and times. Meanwhile, a Joint Working Party of the DfT and the Local Government Association has been considering the implications of criticisms such as those above (expressed by the author and other commentators) for how to monitor progress at a local level. In April this year two short reports were published (DfT/ALG 2004, DfT 2004) giving some indications of the future direction. At present the thinking is that a whole series of different measures of congestion should be monitored, including the 'lost time in comparison with free flow' measure described above, a measure of variability, and a measure of public (or drivers') opinion. The plan is to provide technical guidance on the delay indicator by July 2004: my own view is that this will not be able to identify cost-effective methods of sampling both average and variance, at the local level, sufficiently sensitive to monitor year-by-year progress, under the terms of reference as they stand at the moment. However the direction of movement is the right one and it is reasonable to expect significant improvements in methodology in a short time.

Conclusion

There is wide agreement that congestion is serious, and imposes real costs. That conclusion is robust to method and data. There are also wide perceptions that congestion is getting worse, and will continue to do so unless radical action is taken. That conclusion is, in my view, valid, but it does need careful analysis and thought. Some of the longest-established methods of analysis are not really very helpful in getting to grips with the problem, and although they produce very large numbers, the numbers are not very meaningful. The 'total cost' is a large, but practically meaningless figure, and it would be better not to use it as a target for policy.

It is an unrealistic fantasy to imagine that all road traffic will or can ever be reliably moving close to the maximum speed limit. For practical purposes, transport works much better if the conditions are *not* too close to the maximum that in theory could be obtained. Any system, operating close to its maximum capacity and subject to random variation, is unstable.

Therefore the core of the problem, it is argued here, is that in congested conditions travel times become unpredictable. This, rather than the average speed, ought to be at the heart of policy.

Therefore the two key important things to do are:

- Strategic action to reduce traffic volume to a level where conditions do not vary too much from day to day. Some measures are available which can do this while increasing speed (primarily road pricing), and others which may reduce average speed but iron out the variations and protect specific important classes of traffic (eg priority lanes, signal control). But they both greatly increase reliability, and this is more important;
- Practical measures to provide good alternatives for freight and passenger movements which reduce the intensity of use of scarce road space in congested conditions. Even where this only applies to a minority of movements, significant effects are possible. The main impact will be on selected major routes, and on the companies themselves, and the key test is not what percentage of 'total congestion' this represents (mostly a small number) but how it compares with the cost of implementation.

References

British Road Federation (undated, confirmed as 1988) The cost of congestion

CBI (1989) Transport in London; Task Force Report, London Confederation of British Industry

CBI (undated, 1991?) Paying for Better Motorways: a CBI response

CBI (1993) Paying for Better Motorways: Issues for Discussion

CEBR and Healey and Baker (2001) Impact of congestion charging on London's main offices and retail locations, CEBR.

CIT (1990, 1992) Paying for Progress and Supplementary Report, London.

CfIT (2003) Paying for Road Use, London.

Dargay JM and Goodwin PB (1999) Traffic congestion in Europe, in Road Table Report 110, European Conference of Ministers of Transport, Paris

DETR (2000) (as reinterpreted by Goodwin, 2001)

DETR (2000a) Transport 2010: the Ten Year Transport Plan, Department of the Environment, Transport and the Regions, London

DETR (2000b) Transport 2010 the background analysis, DETR London

DETR (2000c) Modelling using the National Road Traffic Forecasting framework for Tackling Congestion and Pollution and Transport 2010: The Ten Year Plan, Technical Report, DETR London

DfT (2004) Further development of indicators and methodologies, Personal Travel Division.

DfT/ALG (2004) Final report of the working group on delivering success criteria for transport

MoT (1964) Road Pricing: the economic and technical possibilities ('The Smeed Report'), London HMSO.

Glanville WH and Smeed R J (1957) The basic requirements for the roads of Great Britain, Proc. Conf on the Highway needs of Great Britain, Institution of Civil Engineers

House of Commons Transport Committee (1994) Urban Road Pricing, Minutes of Evidence Vol II 104-II 697-I, 1993-4) HMSO

Leach L (2001) More roads and road pricing – the way to go?, IOD Policy Paper, Institute of Directors, London

Maddison D et al (1996) Blueprint 5, London, Earthscan Publications

Mauch SP and Rothengatter W (1995) External effects of transport, Paris, Union Internationale des Chemins de Fer.

McKinnon A (2004) Lorry Road User Charging: A Critique, Logistics Research Centre, Heriot-Watt University, Edinburgh

ROCOL (2000) Road charging options for London: A technical assessment, London, The Stationery Office.

Newbery(2000) in Jenkinson T 'Pricing and Congestion – economic principles relevant to pricing roads, Readings in Microeconomics, OUP.

Royal Commission on Environmental Pollution (1994), Eighteenth Report: Transport and the Environment, London, HMSO

Pearce D W (1993), Blueprint 3: Towards Sustainable Development, London Earthscan Publications

Peirson and Vickerman (1997) in Hensher and Brewer (eds) Transport – an economic and management perspective, 2001, OUP

SACTRA (1999) Transport and the Economy, DETR, The Stationery Office, London

Sansom T, Nash C, Mackie P, Shires J & Watkiss P (2001) Surface Transport Costs and Charges, Great Britain 1998, Institute for Transport Studies, Leeds University.

The Smith Group (1999) Clearing the Way – practical solutions for urban road user charging, London.

Warren J, Potter S, Parkhurst G (2004) Taxation Futures for Sustainable Mobility, Open University/University of West England, forthcoming.