What to do about the railways?

Dieter Helm

22nd September 2015

1. Introduction

Despite the sharp rise in demand and therefore revenue, it is widely agreed that the railways are in a mess. A significant proportion of the electorate think that renationalisation is the answer, and indeed Network Rail itself has been renationalised. Although there have been some successes, the overall performance of Network Rail has been generally poor, and at times appalling. Major electrification projects, like London to Swansea, are behind schedule, there have been problems around London, and the general day-to-day efficiency has been subject to the McNulty Review and constant criticism from ORR.

Matters are not much better elsewhere in the industry. The train operating franchises have also had a mixed record. Some have been persistently bad, many are operated by bus companies, who bring a very different management culture, and service quality has been at best variable. Then there are the other bits – like the train leasing companies, whose profits since privatisation have had questionable tangency with the risks, and which have been more the focus of repeated M&A and big pay offs than on the value for money for rail customers and taxpayers.

This catalogue of concerns is not exhaustive, but it does serve to illustrate a general point: it is very unlikely that tinkering at the margins with incentives will significantly change the outcomes. Furthermore, it is far from obvious that simply nationalising the complete industry will address the fundamental problems that beset this industry. On the contrary, what this catalogue of failures suggests is that something fundamental is going wrong, and that something

fundamental needs to be done. The structure and organisation of the industry in general, and Network Rail in particular, is not fit for purpose.

This paper sets out how this unhappy state of affairs came about, and what to do about it, making the case for a system operator and then the pragmatic disaggregation of Network Rail, with the potential for a wide variety of partnerships and private sector involvements, and gradual re-privatisations. Disaggregation is argued to be both desirable and possible, but only when combined with a strong central system operator, capable of coordinating the system as a whole – within which the parts can then flourish.

2. How did the railways end up in such a mess?

Once upon a time, it was all very simple and straightforward – or so the myth-makers would have us believe. The state owned the railways – from carriages to track to stations. It was one integrated national monopoly, and hence it could be planned. Ministers would decide the priorities, civil servants would monitor the railway's conduct, and the Board of British Rail would pursue a commercial approach with an eye to the public interest, as represented by the non-executive directors – all of whom would be appointed by ministers. Presumably this is what those in favour of the full renationalisation of the industry want to return to.

It was one of the clearest examples of the Morrisonian nationalised industry model, developed in the 1920s and 1930s. For most if not all of British Rail's existence it's task in this Morrisonian structure was managing a decline. In the twentieth century, the car and the truck were replacing the train. Short-haul air travel, especially the budget variety, made matters worse. Transport demand might have risen sharply, but rail demand was falling – relatively and absolutely. Add in the impact on rail freight demand from the decline of the coal industry, and much of heavy manufacturing, and the overall picture was one of gradual contraction, punctuated by bouts of realism, such as the Beeching Report and its cutbacks of the rural and branch lines.

This decline in demand inevitably put the railways under financial stress. A rail network is a *system*, and falling demand does not (Beeching excepted) lead *pari passu* to falls in costs. British Rail was a continuous drain on the Treasury. When privatisation came along, this combination of declining demand and losses was the context. At the time, a rail renaissance was not on the cards. It was all about reducing costs, imposing private sector financial discipline, getting the private sector to finance investment and capital maintenance, and breaking the grip of the unions.

Privatisation followed a twin track approach. First, there was the ambition to create a competitive open access industry – separating tracks from trains, and moving towards a world in which any train operators could bid for any slots. It was a bit like the models developed for electricity and gas. Competition was the mantra, and that meant open access to the tracks. Second, there was the practicality of getting it done, and in ways that would make it as hard as possible to renationalise. It turned out that neither of these would be effected. In particular, the railways were never going to have a full-blooded open access regime, and the way in which Railtrack was sold was seriously flawed – so flawed as to bring it down.

Ministers and officials failed to fully understand that the electricity model was not a blueprint for all industries and, to the extent that it could be carried across to railways, it would need a "pool" to buy and sell track access rights, a system operator to tie it all together, and a financial model which would need to carry the market risks. It could not be done in the case of water, aviation slots were not an open auction pool type model for airports, and no-one could come up with an answer as to how spot competitors could finance trains or provide assurance that the track access charges would be paid.

As a result, a structure for the industry was created (and the one we now have), which was the answer to a resource allocation problem that could not possibly be put into practical effect. It was the wrong answer to what turned out to be the

wrong question. Competition could be marginal, and indeed marginal competition has an important role to play, but there is a world of difference between a bit of competitive open access, and a whole railways system built around this model. Even marginal competition could be problematic.

Leaving the competition bit aside, the practical political issue at privatisation was much more immediate: how to get it done in the face of implacable Labour opposition, in the run up to an election that Labour was expected to win. The Labour Party committed itself to renationalisation – led by Claire Short as the Shadow Transport Minister. This made buying the shares in Railtrack a risky business, and hence they had to be cheap enough to draw in investors, they had to pay high dividends so investors got a payback quickly, and the overall value of the business was therefore set at a low value to reflect the Labour Party risk. Railtrack ended up with a dividend-orientated financial policy, and a very small (£3 billion) balance sheet, unfit for the purposes of carrying a massive investment programme that expanding demand was to require – but then it was declining demand that was assumed at privatisation.

Railtrack ran into trouble almost immediately with the West Coast Mainline upgrade, which ended up costing more than twice its balance sheet value. This in turn meant that it would need more money – either from the taxpayers (directly, or indirectly via track access charges) or from investors (through a rights issue). In the end it needed all of these, and when the Treasury refused to help it get a rights issue away, the then Secretary of State, Stephen Byers pulled the plug, with the regulator Tom Winsor trying to hang onto the legal niceties of the role of independent regulation and his duties to ensure that Railtrack could finance its functions. If there was ever a demonstration of the fact that regulatory independence was relative and not absolute – that it works only as long as politicians would let it - this was it. (That lesson remains extremely pertinent now the government owns Network Rail).

Network Rail was the result. But rather than set up a sensible structure to meet the needs of the industry, a primary concern was that of the Treasury in trying to

prevent Network Rail coming back onto the government's books. In order to do this, a not-for-dividend company was set up, to be governed by "members". The Treasury achieved its objective initially, but at the price of a company with chaotic governance, and indeed so bad that in the end its borrowing would force the result the Treasury had tried so hard to avoid. It would be renationalised. The poor performance noted above was an inevitable result. It is hard to see how it could ever have delivered an efficient outcome.

Network Rail ran up £30billion of debt – debt which it must have known could never be fully repaid. It did this because it had a government guarantee over the debt – magically still keeping it off the governments' books in the short term. Its incentives were perverse: if it could borrow on the guarantee, then it might as well promote lots for capital spending. It could pay its directors what it liked (and it did produce some eye-watering bonuses and pay offs for a company with a public guarantee). On efficiency, why exactly would it push the envelope? Any cost savings would just produce less money as it would result in lower payments from the track access charges via the Train Operating Companies (TOCs), who would in turn get less money from the government. The gradual collapse of Network Rail was the net result of the structure and incentives it had inherited after the collapse of Railtrack. Arguably no set of managers, however brilliant, could have rescued it.

If Railtrack and then Network Rail were both unsustainable, so too was the model for the TOCs. The initial idea was to give out short-term franchises, which would gradually come under pressure from open access competitors - a bit like the early scheme in telecoms at privatisation. With a short term franchise, there were two perverse incentives – to get the highest return in the short run before the franchise expired; and to try to cement in the franchisee so that when it came to re-bidding the incumbent could see off entrants, preferably from even thinking about bidding. The government had thought about the latter a bit and, by having the rolling stock owned by separate companies (ROSCOS), the idea was that if the incumbent lost, the rolling stock would simply be handed over to the entrant.

It soon became apparent that the main entrants were likely to be other TOCs. The re-bidding process was supposed to allow the government to reflect and change what it wanted at periodic intervals, but for the two big franchises – West Coast and East Coast – it turned more into a nightmare. Elsewhere incumbents often survived despite the evidence of poor services.

What had not really occurred to officials at privatisation was that, whereas British Rail could resolve interfaces internally, the new TOCs required well-defined boundaries. It had to be made clear who was responsible for what – *precisely*. Who would run the stations? Who would sell the tickets? Whose fault would delays be? What compensation was to be paid by whom to whom in the event of failures? How would signals and trains interact? Who would be responsible for station upgrades? Rigidity replaced internal flexibility, often with a resulting blame-game when things went wrong.

Finally, the rolling stock companies (ROSCOs) turned into a brilliant way to make enormous sums of money. They were auctioned, and therefore the profits they made could be defended as resulting from a competitive process. But few understood how the railways would work out, and the ROSCOs turned out rather better at making individuals very rich, and providing investment bankers with lots of fees from the repeated takeovers, than they did at making the most efficient contribution to the industry.

3. What are the railways for? What do we want from them?

To be fair to successive governments and to the companies in the industry, the world of continuing decline in demand has been turned on its head. Contrary to the objective of managed decline, and managed decline of the subsidies too, demand for rail has boomed. This is despite significant real price increases. And whilst in other industries there is always the option of either not meeting the demand or charging the full cost of the service, this was never an option in the

railways. Rail fares are, have always been, and will be for the foreseeable future, partly political.

Indeed, it was quite soon after privatisation that the objectives were formally changed. In the 1998 Transport White Paper ("A new deal for transport – better for everyone"), John Prescott set out the bold objective of not just growing rail travel, but of switching on a massive scale from roads to railways. This was to be an integrated transport strategy. He famously said that he should be judged on this switch (and since he massively failed to do so, one might have expected him to resign from government – which, of course, he did not).

The game was now very different – and with very different costs. It would require a U-turn from the Treasury. Instead of phasing out the subsidy to the railways, now it would be permanent. The fact the Prescott was a "big beast" in the Labour Party made it easier to push the Chancellor Gordon Brown to accept this. Yet the old problem that had beset British Rail had not been solved: how exactly could the Treasury credibly commit to medium to long-term funding in the context of short-term budgetary considerations? The problem remains today.

To effect this new rail strategy, Prescott set up the Strategic Rail Authority (SRA) as an enhanced franchising office. Alistair Morton as appointed to run this, and he saw his role as planning the future shape and development of the railways. We now had the SRA and the Treasury and the Rail Regulator (ORR) all vying for control and in a crowded space of companies – Railtrack, the TOCs and the ROSCOs, plus the rail freight companies.

The trouble with this set of confused and overlapping actors is that there was no clear translation of objectives into strategies, investments and outcomes. The SRA did not solve any of the key problems, and it would not survive the demise of Railtrack. This can be read two ways. Charitably, things just did not work out as intended. Alternatively, the Treasury never signed up to the objectives in the first place – Brown was playing a game with Prescott. Let him promise the radical growth and switch from the roads, and then starve the industry of the

extra funds this would have cost. Sadly, the latter is the more convincing explanation.

From 2000 onwards, the conclusion that emerges is that the industry has operated in a context where *there are no clear overarching objectives*. The questions - What are the railways for? What do we want from them? – have been given no answers. Instead the railways have muddled along, treating each problem incrementally and largely in isolation. Thus with each franchise re-letting the Department for Transport (DfT) sets out its particular wish list, the Treasury considers the implications, and then the companies bid. For Railtrack and then Network Rail, the company negotiates with ORR on its capital programme and capital maintenance, and then there is a heated argument about cost efficiencies.

The Treasury temptation in this confused context is to exploit the leeway it gives. With a Chancellor who likes big projects, it is HS2 and electrification of the mainline from Paddington that get the go-ahead, even if this means squeezing the monies for the more boring day-to-day capital maintenance and small scale enhancements. As the financial constraints bite, some projects get cut. Railway policy becomes a list of such projects – rather in keeping with the project list approach that now dominates the overarching National Infrastructure Plan.

All this highlights that the major problem in trying to design the structure of the railways is that there are no clear objectives. The government has not spelled out what it wants from the railways. It has not said what it thinks the railways are *for*.

4. Three options for the way forward

What sort of answer could be provided? There are three broad options. Option one is to go back to the aims of 1980s and 1990s: to make the existing railways less reliant on subsidies, limiting investments and expansion. Option two is to re-visit Prescott's White Paper and target an expanding railway for the next few

decades ahead whilst retaining the existing structures. Option three is to restructure an expanding railway industry in an evolutionary way, taking account of rising demand, and providing a coherent coordination of the gradually more disaggregated parts. These might be regarded respectively as: retrenchment; trying to put new wine in old bottles; and a pragmatic reform programme.

Option one is not quite as redundant as it might seem. The existing structures have proved capable of transporting many more people than most had anticipated (though in the 1920s they carried a lot too). Managed decline might simply mean that the railways are forced to make a profit, *but not expand*. This has two implications: the existing network has to be used more intensively – to get more revenue from each asset, given the marginal cost is below the average – and prices have to go up to reflect the full costs. The former could be somewhat modified by allowing for IT to be applied. Modern smart signalling and judicious investments in greater platforms could move the main rail networks towards the sorts of frequencies that are witnessed on the London Underground. Anyone waiting for trains at mainline stations even in rush hours will have been struck by the fact that, unlike the roads, the track is mostly empty.

The case for sweating the existing assets – or at least limiting major expansion – as the objective rests on analysis not just of railways, but of transport more generally. The reason trains have declined since the Second World War is because cars, buses and trucks have proved much more attractive ways of travelling. It is far from obvious that the future is with trains, and that the Prescott objective of getting people out of cars and trucks is a good idea. But if it is to be cars and trucks then the policy implication is that the priority is to do something about the roads and their use. Some obvious considerations are: driverless cars, electric cars, smart traffic management and dedicated bus lanes. It is a reasonable question to ask as to whether trains are really better than buses for urban and commuting travel. If buses had the sort of dedicated space granted to trains, their flexibility would be in contrast to the key feature of

railways – their inflexibility. Furthermore, with electric vehicles, the environmental impacts would be rather different.

Option two is an expansion largely within the current structures – a version of predict and provide. That is the objective when it comes to high-speed railways. HS2 is a £50 billion project to create a massive piece of infrastructure – possibly the biggest single project in Britain. Yet even here the objective is not clear, as witnessed by the confused way in which the government has sought to demonstrate its benefits. What is the question to which HS2 is supposed to be the answer? It might be thought that it is to join up the high speed rail system to HS1 and to the continent of Europe. But no – the one cost saving in the £50 billion that has been identified is the connecting up of HS2 with HS1. £0.5 billion is the estimated saving of making passengers get out at Euston and then work out how to get to St Pancras – leaving the other £49.5 billion to be spent. So the objective is no longer an interconnected high-speed rail system. It must therefore be primarily about increasing capacity between London, Birmingham and the North, and then it is to be compared with other ways of achieving this capacity enhancement.

What would this expanded option two look like? What might the rail passengers expect? The fundamental questions are usually ducked. Will this be a punctual service - a critical systems choice? Will it be high speed or slower but with less variance in arrival time? Will it add capacity ahead of demand or will it be permanently overcrowded? Ultimately, will the government predict correctly, and will it be prepared to provide?

This is where *option three* comes in. Option two requires the ability to predict demand – and then provide it. Option three recognises that demand is disaggregated into lots of sub-markets, and that different objectives might be applied to different bits. This may seem obvious, but it has major implications for the structure of the railways. There are a variety of ways of disaggregating, and given the system interactions, there is no perfect way to do this. Yet it is possible and practical to separate out city transport systems, inter-urban transport,

commuter routes, and regions. Each has its own competitor transport systems – from walking and cycling to buses in urban areas, to short-haul flights for inter-urban transport. Option three asks the question: what do we want from *each bit*?

Considered in this way, the case for a national network monopoly looks less than convincing. It is not obvious why Network Rail should operate HS1 or HS2 and it does not in the first case and will not in the second. Similarly for London. The Great Western franchise is distinct and to a considerable extent discrete. The questions then are: do we want a high-speed international network? Should Great Western by electrified and what connectivity in this region is optimal? How should London and Manchester coordinate their multiple systems? It is also far from obvious that national standards and national measurement of performance are appropriate. Commuters might place much greater emphasis on precision timing – though not if there are trains every 10 to 15 minutes. In rural areas, the choice between bus and rail services is one in which the flexibility of buses might trump rail – assuming that they are actually provided.

All three options can be defined in economic efficiency terms. But a particular feature of core networks is that they have a social dimension. Access to transport, electricity, water and now broadband are necessary to participate in society. In the peripheries of systems, not even the marginal costs of providing rail services can be covered. The question then is: how much should transport be subsidised? And if it is, how much of this should go to buses, and how much to rail? The critical point here is that the social consideration is about transport, not about any specific form.

The final dimension that all three options throw up is about spatial planning and development. Each of the options changes the specific locational attractions of housing and business developments. We could, for example, take the view that if it is expensive to provide more rural transport, then people who need to travel should simply move out of the countryside. They should move to cities, as they are around the world. It is much easier to provide transport, electricity and broadband in cities. Let them see the true economic costs of living in low density

areas, and let them decide whether to bear the costs. Alternatively, we could encourage rural developments. Similarly, HS2 can be built to connect up existing industrial, business and residential areas – reinforcing the current spatial configurations – or it could deliberately go to other areas, on the assumption that business and houses will follow. It could be a means to regional development or it could reinforce the existing patterns – and draw yet more people to London.

Spatial issues are about *systems* too. Each decision imposes costs and benefits on everyone else, and each shapes society in particular ways. The rail networks both cause particular spatial distributions, and are a response to existing distributions.

The problem with all these possible options is that they need *co-ordination*, though less so in option one. If the railways are left to fend for themselves, they can be taken as largely given. Options two and three require a great deal of coordination. *Someone needs to be in charge, to make sure the system as a whole is considered*. This is the role of the system operator, to which we now turn.

5. Who is in charge? The case for a system operator.

Network infrastructures raise special economic problems. They are more than the sum of a set of discrete assets and investment projects. Change one bit of the system and it effects every other bit. There are often massive externalities from any specific decision. At the most basic level, timetables need to be coordinated – between trains, and between trains and other modes of transport. This is essentially the short-term coordination problem. But investment too needs to be coordinated. Building HS2 will impact on all the existing rail networks, and in particular the East Coast Mainline. But it will also impact on the London Underground, depending on whether the trains arrive in Euston or join up with HS1 at St Pancras. It will effect the economic case for a new runway at Heathrow. It will also impact on air schedules and on motorway provision. This is the medium to long-term dynamic coordination problem.

All network infrastructures face this top-down coordination problem. Some do this explicitly, like electricity transmission and distribution. Others internalise within network monopolies, like water and railways. Some don't coordinate the systems very well – like mobile phone networks and broadband.

The electricity example does provide one insightful lesson for the railways. The static problem in electricity is that at each moment in time demand must be matched with supply – otherwise there are blackouts. Electricity cannot be stored, except in very specific circumstances. The power stations need to be dispatched in such a way that they add up to total demand – which varies through the day and night considerably. There can be surges in demand (for example during the adverts in key TV programmes) and there can be unanticipated failures to supply (for example, a power station failure or a transmission line problem).

To meet the need to keep the lights on, the electricity system needs a system operator, whose job it is to schedule plants on and off the system in response to demand, and to ensure that there is enough spare capacity in the system to deal with unanticipated demand surges or supply failures. It needs to coordinate the supply of electricity and to ensure there is a sufficient capacity margin. It does not need to own any power stations, but they must do what they are told by the system operator. Nor does it need to run the transmission and distribution lines – this is for the transmission operators and the distribution companies.

The analogy with the railways is obvious. The trains cannot be "stored" and there does need to be an excess capacity margin to deal with shocks – to the track (analogous to the transmission and distribution lines) and the trains (analogous to power station failures). Yet unlike the electricity example, if we ask "who is the system operator in the railways?," the answer is far from obvious. Network Rail is in charge of the signalling and the track and it has an overriding command-and-control function. But it is not separate, even in the short run. The system operator role in respect of the timetable is also vague. There is no posted genuinely national timetable as can be found for example on the Swiss railways.

There is also the interface with the regulator and government. The investments – and thus the capacity margin – are determined in a tripartite way between Network Rail, ORR and DfT (or perhaps more accurately by the Treasury). It might be argued that the whole point of the DfT is to provide a coherent overall view of the transport system and how the various parts fit together – a dynamic coordinated picture. For much of the post Second World War period, this is what its various incarnations were supposed to do. But privatisation and, in the case of buses, deregulation put paid to this in the 1990s, and with inconsistency between the charging for the rail track and the roads, and the differential taxation regimes, the incentives have probably made the misallocation between the transport modes, as well as within them, gradually worse. Imagine, for example, if the DfT had seriously compared a new (high speed) motorway, an upgrade of the existing rail and road system, a low speed freight dedicated motorway, and a dedicated long distance bus road system *before* going for the HS2. More parochially, consider the choice between a new bus-only motorway lane from Bristol to London, versus electrification of the Great Western line.

It can at least be said for the 1998 White Paper that it got the overall question right. Whether the answer was right is a matter of working through the consequences of implied massive switch to railways that it required, providing the finance and delivering the investment. As noted above, the Treasury was never really on side, and so the objective was at best irrelevant and at worst likely to produce considerable waste and inefficiency - which it did.

The Prescott strategy did have one further positive feature, even if it too was very poorly executed. The SRA was set up with many of the features, which a system operator might have. Yet instead of giving the SRA responsibility over the industry, it was stymied from the outset by a combination of its inheritance of rail franchising from OPRAF, and the role of the competitive institution ORR, and again the Treasury's priorities of reducing the dependence on public funding.

The role of ORR, like that of the other utility regulators, started out as essentially an auditing one. Its job was not to decide what could be done (the outputs) nor how the railways should be financed, but rather to make sure that the costs of delivering the outputs would be efficient. But it never turned out like this. The regulator had a role in protecting customers and protecting investors, and therefore was directly involved in prices, resource allocation and the ability to pay, and also in finance, since only then could it make sure the companies could finance their functions.

To this inevitable wider role, ORR like all the other regulators has been involved in "mission creep" and not entirely at its own volition. Politicians look at outcomes, and put pressure of regulators to "do something". Regulators have been gradually given wider and wider remits across the utilities. They end up involved in environmental matters, safety regulation and even salaries.

In the railways, this crowded set of overlaps has become a serous impediment to the efficient running of the system, and has increased costs and uncertainties. There is no overarching transport policy, there is no clarity as to the roles and functions of the Treasury, Parliament (now Network Rail is nationalised), the DfT, ORR, Network Rail, the train operators and increasingly local authorities, the cities, devolved administrations and their respective responsibilities. Further devolution of transport to the Northern Powerhouse, London, Cornwall and probably a host of other cites and regions will further exacerbate the overlaps. *The more the system is broken up, the greater the need for a central system operator* to define the relationships and powers in respect of all the other parties.

6. How should the railways be paid for?

Unless the government seriously wants to pursue option one, the railways are unlikely ever to make a profit in aggregate. HS1 has been sold off, and HS2 will possibly be, but the fact that they are built in the public sector tells us that the private sector would not build them. The commuter and other rail fares

necessary to break-even are politically impossible on a full cost basis. The conclusion is obvious: *subsidy is a permanent feature of the railways*.

A second key financial consideration is the efficiency one. The marginal cost (except at congestion points) is very low relative to the average cost for both road and rail travel. An empty seat on a train and an empty motorway has zero marginal costs. If one person uses either, they add nothing to costs and they gain personal benefit, and hence it is sensible to price up to the congestion point at just above the marginal cost so they make some contribution towards the system fixed and sunk costs.

The problem is how to tax the customer base to cover the additional fixed costs over and above the marginal costs. Here the subsidy issue needs to be considered in a wider context. Transport as a whole (as opposed to particular modes and services) is extremely inelastic in demand. Indeed that is why it has been a target for taxation, notably for petrol and diesel. We should prefer to tax those things that are inelastic in demand, since the tax will not distort choices.

Dig a little deeper and it is obvious that within each mode, some uses are more inelastic than others. This is the classic argument for peak pricing. Commuters have to get to work on time, children have to get to school on time, but pensioners and shoppers can often choose when to travel. The economically efficient solution is to charge a lot to commuters and to let the school run ration by congestion.

In aggregate, the government receives the petrol taxes, VAT, the taxes on car purchases, and the vehicle road fund licence. It provides the capital for road building and road maintenance (the latter partly through local government spending), and the subsidies to the train operators and to Network Rail. It also indirectly contributes to bus services and city transport systems. There is no reason why the two should not add up – *the money raised should and can equal the total expenditures.* The inelasticity of transport demand makes this entirely feasible – transport *as a whole* can pay for itself *as a whole.*

The financing problem is then a resource allocation one – how to allocate the income to the specific and competing expenditures. Should the government find £15 billion for the strategic roads and £50 billion for HS2? Should it prioritise other transport areas? Or should it give a block transport grant to cities and local authorities and leave them to sort out the consequences?

It is immediately apparent that none of these questions can be answered without first sorting out the objectives. The allocation follows from the objectives, and given that the objectives are not clearly specified the result is the risk of a lurch from one pet project to another, and a response to each "political crisis" transport throws up.

7. How should the railways be structured?

Structure, like finance, follows from the objectives. But it also follows from the functions and the underlying cost characteristics of the industry. Since privatisation, the structure has in fact seen considerable evolution – from short to longer-term franchises, from private risk to government guarantees, and towards widespread devolution to the various layers of government. The current proposals for devolution to the big cities and to Cornwall are more radical and suggest a more substantial shake up.

Renationalisation of Network Rail and the problems of the East and West Coast mainline franchises have created a further opportunity. The government now has the ability to decide what to do with Network Rail because it controls it – in a way it could not do so easily with the not-for-dividend company (indeed Network Rail was designed to limit government control, in order to keep it out of the government's accounts).

The first task is to decide what the objectives are. Only the government can do this, and it should.

The second task is to set up a transport account for the whole sector – recording monies in from the various taxes and charges, set against the current and capital expenditures. The government needs a *national transport balance sheet*.

The third task is to carve out a clearly defined system operator function from Network Rail, and specify its precise roles and functions, and its powers to direct at the national level. The greater the disaggregation, the greater the need for a national system operator. This needs to follow the *deep* SO model, and in the process it can assume a number of investments functions from both DFT and ORR – leaving the former to concentrate on the overarching policy formulation and the latter on the review of efficiency dimensions.

With the system operator in place at the national level and hence ensuring the national coordination and the necessary system planning of the overall system is not lost, there is no good reason for keeping the rest of Network Rail intact. It has not proved itself to be efficient (indeed quite the contrary), and there is much to be said for having a number of comparator parts rather than a monopolistic block.

Break-up should follow the discrete cost structures within areas and services, and take account of the extent to which the rail, bus, and local transport are substitutes. It should be pragmatic and evolutionary, rather than adhering to a top-down blueprint.

In breaking up Network Rail, some parts are more obvious to define than others. The clearest examples are:

- *the HS networks*. As new and separate infrastructures, these can be run as separate services, as indeed HS1 is already. It is an open question as to whether HS1 and HS2 should eventually be merged – and this turns in part on whether they are actually going to be joined up. But HS2 is a development project, and it will be years before the issue of merger

arises. Nevertheless they should fall under the overarching remit of the system operator.

- *the big urban centres* have the greatest interactions between modes of transport and in respect of planning and overall coordination of local government functions. The process of devolution has already begun, but largely for business and urban transport systems. The practical question is where the economic boundaries of cities stop – or rather where their impacts begin to peter out. In the case of London, commuting adds a couple for million people per day, and there is a good case for widening the remit of TfL to incorporate more of the surface rail networks, and perhaps even to take control of the main London stations. Control of stations at the urban level has considerable advantages for the Northern Powerhouse and other major cities.

- *Rural devolution* at the periphery has considerable advantages, since the problems of bus versus rail, and local road provisions typically have a well-defined spatial context, and because they are typically at the end of the transport networks for both rail and strategic roads. Rural land use planning, rural social services and rural economic development are very much local planning matters. Since subsidy here has an inevitable role and is probably permanent, a more focused approach to value-for-money would follow, and it would enable the separation out of the more profitable elements of the networks.

- *Some regions have well defined and largely separate networks*. The railway industry grew up as a regionalised oligopoly with the interconnections focussed around the big cities, and overwhelmingly London. Thus Great Western is fairly discrete. Decisions about how to develop and manage Reading, Bath, Bristol, Swansea, Exeter and Penzance stations and services are all about the regional systems.

- *The intercity long distance routes are fairly discrete*. There could be a north-eastern and a north-western integrated regional company, or narrower northeast and northwest intercity companies split out. These could follow the lines of the pre Second World War companies (as with Great Western) or be more tightly focussed.

This leaves out full vertical integration, which is what what the unions and some of the train operators have continued to argue for, and what would be the effect of renationalising all the industry and thus having a single owner – the government. If only the track and trains were put back together, it has been claimed, all would be well. The problem with this is twofold: all was not well in the vertically integrated monopoly; and the coordination problems that vertical integration is supposed to solve can in any event be dealt with by the system operator.

There is a case for a mixed approach – some vertical integration where there are obvious gains – such as in cities and in some regional networks – and some separation where the functions remain clearly distinct. EU directives in any event limit the options - some vertical unbundling remains a requirement.

8. Does it have to be a big bang, or can there be an evolutionary path?

The obvious problem with the piecemeal break-up of Network Rail proposed above is that it is messy. Instead of the neat certainties of British Rail, or the privatisation structure where each part had its particular roles and functions, the disaggregation along urban, regional and local lines does not have such clear lines in the sands.

In reality, such lines in the sand have never been quite that clear. London has been different from the outset for example. HS1 is another. But even if they had been, they did not reflect the costs and competitor systems on the ground. Substitution really is different in cities, and in regions, and on long distance routes.

There is a good case for plurality. It is not just recognising the different circumstances, but also the introduction of competing business models and management styles. Given that Railtrack and then Network Rail have never managed to meet the efficiency tests that ORR have repeated conducted, given the mess over recent management of upgrades, Christmas works overrun and the electrification of Great Western, and after nearly two decades of experience, it is reasonable to conclude that the current structure of Network Rail is never going to be efficient. No short-term changes in management will make much difference. It is obvious that the incentives – and the structure of Network Rail- are not fit-for-purpose.

Plurality has one further benefit. It can be done gradually in an evolutionary way. A series of incremental steps can be taken – and indeed some are already in process. London can be granted a greater role in surface commuter services. The government could transfer the key stations incrementally to TfL – it does after all own both companies. Cornwall's powers can be enhanced, and the Treasury provides the money in any event so it can change the allocation rules. The private sector can play its part on a case-by-case basis.

Network Rail would, in loosing these bits, remain to handle the bits left. There is no inherent reason why with less to do, it could not run the remaining bits better. It would, after all, be hard to run them worse than at present. But clarity would be better provided if its future were set out. Whilst holding the fort in an evolutionary way as parts for the network are broken out to the cities, regionals and local areas, there is a considerable difference between a world in which the objective is to close it down eventually (leaving the system operator in place) and one in which it remains one player amongst many. The former is neat, but unlikely and probably unnecessary.

In this regard, the gas sector is interesting as a parallel. British Gas, and then Transco, once ran all the regional gas companies. Now National Grid (which took over Transco) runs some, with others in separate hands. It has not been divested

of all its regions, and by having benchmark competitors its management has been more focused and efficiency has improved. A slimmed down rail track company would have an analogous role, and there is no reason why it would not perform better given the spur of the management competition from the others.

With the separation out of the system operator and the gradual disaggregation, Network Rail could be re-privatised. The evolutionary approach would start with the divestment of specific assets following the restructuring set-out in the previous section above – passing control over some stations, urban networks and regional systems to new and existing entities. This is much easier to achieve if the government owns and controls Network Rail. It is doubly so once the debt sits with Network Rail, underwritten by the Treasury. Once the disaggregation has been completed, then the rump company can be sold to the private sector as a utility, with the usual utility characteristics.

In this evolutionary transition, the subsidies will still have to flow. There will however now be a new context: London, Manchester, Cornwall, Great Western, North West and North East and so on. It will no longer be a single block of money going indirectly or directly into the black hole of Network Rail. This should make it easier to sort out – and earmark – what the subsidies are actually for, rather than leaving this to be worked out internally in Network Rail.

In all this process the glue that holds the evolutionary process together is the system operator. This has considerable power and it has public responsibilities. This should be kept in the public sector for the foreseeable future – but at arms' length from DfT and Treasury. In creating and underpinning the system operator, the functions of the DfT and ORR will change. DfT will be more focussed on objectives rather than detailed capacity decisions and coordination, and ORR will be able to retreat from the wider ambit it has been necessary for it to develop. Thus the multiple government and regulating interactions – and the costs that come with the overlaps – should go down.

9. Conclusions

Sorting out the railways is not an intractable problem. But it does require clarity about the objectives – what the railways are *for*. Rather than "patch-and-mend", treating each problem discretely, a better way forward is to decide what we want for the railways.

The current model is not working, and after a couple of decades of experience, it is obvious that it is not going to work without some major surgery. That can however be achieved in an evolutionary way. Network Rail in its current form has failed, and a partial divestment of some of its assets to cities and to regional entities provides a controlled way forward – provided the system as a whole is glued together by a system operator.