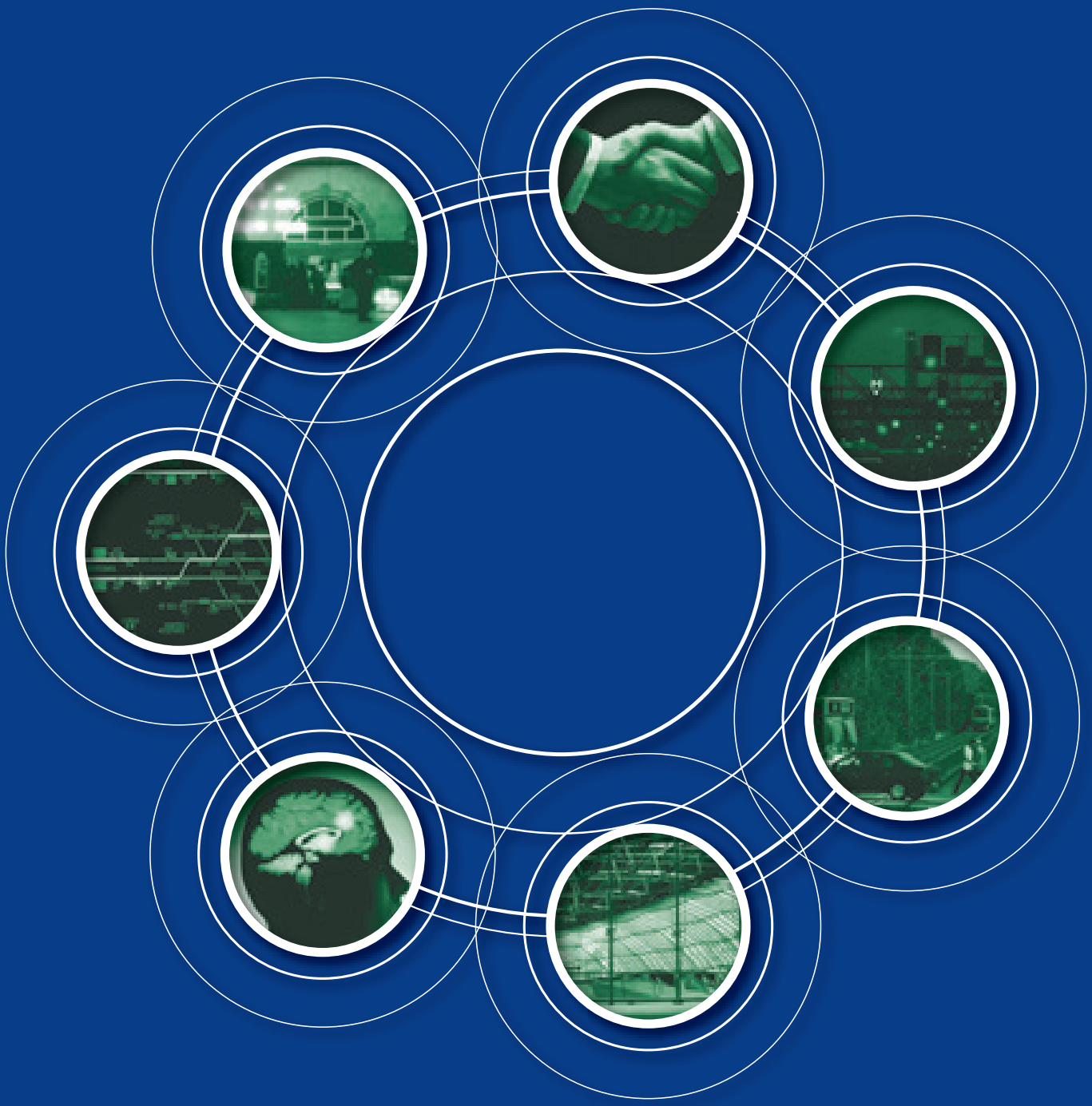
Rail Safety & Standards Board

Research Programme

Management

Railway safety and the ethics of the tolerability of risk



Railway Safety and

The Ethics of the Tolerability

of Risk

Study commissioned by RSSB

Jonathan Wolff
WCA Consulting
3 Calais Street
London SE5 9LP
jwolff@wca.co.uk

October 29th, 2002

Contents:

Executive Summary

1. The problem addressed in this report is that on the one hand the railways are perceived to provide a very safe form of transport, but on the other hand the public appear to want vast sums spent on further safety improvements. The task, thus, is to see what can explain this unusual divergence in attitude, and particularly whether it is driven by ethical considerations. A further question is whether anything may be done to reduce this divergence. (Part 1)

2. The Health and Safety Executive Document *Reducing Risks, Protecting People* provides much helpful background material. However the notion of societal concern requires more sustained conceptual and moral analysis than is provided by HSE, and the HSE approach can do little to advance the analysis needed here. (Part 2.1)

3. Psychological factors, such as social amplification, signal value, quasi-elimination and the value of a loss can all play a part in the analysis of attitudes to railway safety. Nevertheless it is suggested that the idea of preventable systems failures – both the failure of an existing system and the failure to introduce a system - is key to understanding the initial problem. (Part 2.2)

4. Standard risk cost-benefit analysis is based on the assumption that the acceptance of risk in an individual's life is a guide to that individual's broader social tolerance of risk. It is argued that this assumption is not always correct, for it fails to distinguish risk from the imposition of risk; a distinction which is needed for a more developed moral and analytic conception of societal concern. When it is felt that risks are being imposed in ways that are morally problematic a form of societal concern may arise which is not based on any risk assessment, but on moral criticism of those who are seen as imposing the risk. People can become deeply concerned that certain behaviour (perceived to be immoral) exists, and even ashamed that their government tolerates such

risk imposition, even if those individuals show low concern for the risk itself. This appears to be the case with railway safety at the present time, in the view of some members of the public. (Parts 3.1 – 3.4)

5. Owing to intense media scrutiny, the industry suffers from social amplification effects; however this is not the standard case of amplification of risk, but amplification of moral culpability, where the belief in culpability is not based on a concern about individual risk. (Part 3.5)

6. It appears that the core of the problem is that the public will not accept that the industry is fulfilling its safety obligations unless it adopts a state-of-the-art approach to railway safety. Hence the public do not believe that currently railway safety is dealt with properly. It is this that drives the moral criticism of the industry, which, in turn generates immense social concern. (Part 3.6)

7. The implications of the analysis are that, first, a more consistent way of taking public attitudes into railway safety decision-making needs to be devised, and second, public attitudes cannot be addressed and challenged until there is a better understanding of the root of the demand for a state-of-the-art approach. (Part 3.7)

8. Several further projects are proposed: a psychological study to understand the assumptions which underlie public preference for a state-of-the-art approach; an investigation of legal and philosophical understandings of culpable negligence; a public relations exercise both to correct for false assumptions and to do more to explain railway safety improvements and policy to the public; and a revision to the decision framework to take explicit account of commercial factors; and an examination of the possibility of providing a new framework which goes beyond cost-benefit analysis. (Part 4)

Part 1. The Central Problem

This project starts from a conundrum about public attitudes to railway risk and safety. On the one hand studies indicate that, if anything, the public are less concerned about risks from railway travel than they are about risks from road travel, and the conclusion from such studies is that relatively little additional money should be spent on safety improvements. [Jones-Lee, 2000a] Yet on the other hand, the very same members of the public will argue that the fact that certain accidents have taken place is completely unacceptable, and that the managers who allowed such things to happen border on the criminally negligent. [Jones-Lee, 2000b] The question addressed by this report is what can explain this divergence, and, specifically, whether it is driven by ethical assumptions. [Sharpe, 2002] This leads to further questions about what can be done to reduce this extreme response, and how these attitudes should influence railway safety decision making.

The background assumptions are, then, twofold. First, that in general people think, and act as if they think, that traveling on the railways is a very safe form of transport. Few people will refuse to take a train for safety reasons, modify their behaviour to take account of the risk of a train journey, or even think about safety issues when they step on a train. Second, many members of the public believe that very expensive further measures should be taken to improve safety, especially with respect to avoiding systems failure.

It is worth stating at the outset that this is a relatively unusual problem in risk analysis. The risk analysis literature contains the analysis of many cases where there are conundrums about risk. [Slovic, 2000a] However most of these concern cases where expert risk assessment is out of step with public perception (e.g. exposure to nuclear waste and other radiation risks [Slovic, 2000b]) where people modify their behaviour, or fail to do so, in ways in which risk experts consider to be irrational, or where sociological groups make differing risk assessments (e.g. educated, right-leaning, white males typically assess risks as lower than other groups. [Slovic, 2000c]) I have not, however, seen any reference to any study of a case where there appears to be such divergence in

individual attitudes to safety risk in that the same people appear to occupy such conflicting standpoints. While we will see in Part 2 that there are factors – specifically framing factors – which can induce individuals to have highly inconsistent attitudes to the same risk, the divergence experienced with respect to railway safety goes beyond what can be explained within the existing framework.

In what follows, first the relevant literature will be reviewed. Then an analysis of the problem will be proposed which uses aspects of the current literature but goes beyond it in some respects in proposing some further conceptual and moral analysis. The implications of this analysis will be examined and prospects for future action considered.

Part 2. Background

In Part 2 some of the relevant literature is reviewed; in particular the Health and Safety Executive (HSE) guidelines, and some of the psychological literature. The former introduces some useful terminology, but also has some important conceptual limitations. The latter contains a sophisticated analysis of individual perception of risk. However its scope is too narrow to provide a complete solution to the initial conundrum, even though it makes a useful contribution.

2.1 HSE Guidelines

The most important of the Health and Safety documents is *Reducing Risks, Protecting People* [R2P2]. In the first instance this comes across as a sophisticated and well-informed account of the issues, making many helpful points and distinctions. Yet despite the fact that it mentions railway safety several times, it may be hard to apply to railway safety. It may also have some limitations which will be mentioned below (see section 2.1.2 below and Part 3). Nevertheless there are a number of concepts presented and distinctions made in R2P2 which will be helpful for the following discussion.

2.1.1. Hazard and Risk

Roughly speaking a hazard is a feared event and the risk its probability of occurring and causing harm. This is a straightforward enough distinction, but useful for making clear that sometimes a hazard is so awful that no risk of it, however small, is acceptable.

2.1.2 Individual and societal concerns

This is a very important distinction in the current context. Individual concern is defined as ‘how individuals see the risk from a particular hazard affecting them and the things they value

personally.' (R2P2, p. 12) Societal concerns are 'risks or threats from hazards which impact on society and which, if realised, could have adverse repercussions for the institutions responsible for putting in place the provisions and arrangements for protecting people .... Societal concerns [sic] due to the occurrence of multiple fatalities is known as societal risk.' (R2P2, p. 12)

This distinction is something of a puzzle. It is hard to see exactly what the notion of societal concern is intended to capture, or why societal concern should exceed the sum total of individual concern perhaps with an added 'altruism' value which reflects people's concern for each other. (Note, though, that this may be included already under individual concern, as defined, although this is somewhat unclear). But why should societal concern go beyond individual and altruistic concern?

One possible answer is that where a hazard is of a particularly devastating nature, but its probability of occurring is very low, or very unclear, standard calculations may lead to the conclusion that the risk is negligible. Yet people will feel that special precautions are necessary because of the nature of the hazard, however unlikely. For example, the commissioning of a 'cheap' nuclear power station is unlikely to be acceptable if it has a chance of causing greater loss of life than a more expensive design, even if statistically the chances of dying remain virtually nil.

However this does not seem to be an adequate explanation. Although it explains why statistics may not always be a good guide to risk, the factors that make this so should influence individual attitudes to risk too. It fails to explain why a notion of societal concern is needed over and above individual concern.

Furthermore it is unclear whether 'societal concern' is meant to be 'concern of society about risks to individuals' - that society itself, over and above the individuals who make it up can have concerns - or 'concern individuals have about society' as in 'what is society coming to if things like this are happening?'. Indeed, it could even be 'concern of society about society'.

It appears possible that all of these ideas, which are quite different, are included under this single notion within R2P2, and that the notion of societal concern is not always used in a consistent fashion throughout the literature. It seems also sometimes to be run together with the idea of altruistic concern, which, as defined, could be part of individual concern.

Furthermore, the notion of 'societal risk' is defined simply to mean 'societal concerns due to the occurrence of multiple fatalities in a single event', which seems a confusion. Society could be both concerned about, and feel threatened by, risks that might cause only isolated fatalities, or not cause fatalities at all, but, say, make very many people homeless.

It can be conjectured that the notion of societal risk was initially introduced in the context of nuclear risk, where the potential for catastrophe far exceeded previous man-made hazards. Arguably this engendered a new level of risk and concern, and the term 'societal risk' was coined to capture this hazard. However, since then the term has become extended in its usage, and, it appears, has suffered a loss of clarity as a result.

I will return to the idea of societal risk in more detail below. My suspicion is that the idea of societal concern as it is now used in R2P2 and the following literature presupposes a richer conceptual and moral framework than that set out in the report. Part 4 below attempts to provide a more comprehensive conceptual and moral analysis of societal concern, which will assist in the analysis of the central problem.

2.1.3 Unacceptable, tolerable, and broadly acceptable risks

The HSE make a useful distinction between unacceptable, tolerable and broadly acceptable risks.

The basic principle is that where a risk falls into the intolerable region (defined in terms of probability of death per person exposed) then it simply should be prohibited if the risks cannot be reduced, unless exceptional reasons apply. If the risk falls into the tolerable area then the principle of making the risks 'as low as reasonably practicable' (the ALARP principle) is applied. This is a form of cost-benefit analysis (CBA) where the costs and benefits of safety improvements are compared. The benefit of saving a statistical life is translated into financial terms for the purpose of this analysis, and a standard figure is used for the value of preventing a (statistical) fatality (VPF) which is currently around one million pounds in many contexts. If a safety improvement can reasonably be expected to save x lives and cost less than x million pounds, then according to the ALARP principle it should be adopted. Indeed any safety measure should be adopted unless the costs are 'grossly disproportionate' to the benefits. 'Gross disproportion' is not defined, although there seems to be some suggestion that the closer the risk gets to the intolerable level the more should be spent. This implicitly assumes a 'tolerability' multiplier, which is explicitly adopted in Marris [Marris 1993, p. 23] although apparently not elsewhere.

However R2P2 cautions against simplistic use of cost-benefit analysis. In effect it argues that although where there is intolerable risk the ALARP principle does not apply ('an individual risk of death of one in a thousand per annum should on its own represent the dividing line' (p. 46)) there can also be cases which fall within the definition of tolerable risk, but, still, it would be wrong to use the ALARP principle. So there could be other reasons for banning an activity, or requiring a higher level of safety, even in cases which fall under the definition of tolerable risk. Thus it also notes that 'these limits rarely bite'. (p. 46) That is, anything that is an intolerable risk by this standard is also likely to be intolerable for other reasons, and this may also apply to risks that clearly fall within the tolerable level. So in effect it also operates with a concept of intolerable

hazard, where 'the potential severity of the consequences, rather than the probability of them occurring, is the dominant consideration' (p. 27). It suggests that the following types of hazard need special treatment:

a) those which could lead to catastrophic consequences
b) where the consequences may be irreversible
c) which lead to inequalities
d) could pose a threat to future generations

It has been suggested that the first of these applies to railway safety, because of the possibility of multiple-fatality accidents. [Sharpe, 2001] However it has also been questioned whether this is so [Marris 1993]. It is hard to know what is to count as a catastrophe, but seven deaths – or even 30 – in a railway crash would not normally be considered such. Recall that these principles were originally designed in the context of the nuclear power industry. I would imagine that this is intended to cover incidents that may cause hundreds or even thousands of deaths, but where the risk is so small that a figure of 'expected deaths per year' is highly misleading.

Nevertheless it will be suggested in the following that there are reasons for treating certain multiple-fatality accidents in a special way.

2.1.4 Equity, utility, and state-of-the-art

R2P2 points out that three approaches to regulating risk are commonly used:

- equity which is concerned to ensure that everyone's right to protection is guaranteed;
- utility which tries to balance costs and benefits;
- 'state-of-the-art' which suggests that all known techniques should be used.

The document also points out the difficulties with each of them – equity requires too much attention to worst-case scenarios; utility ignores ethical considerations concerning the distribution of risks; and state-of-the-art is very expensive. It argues that the Health and Safety Executive's own approach is able to balance the considerations against each other. One point of interest though is that the central problem can be expressed in this language: the very same people will apply utility and state-of-the-art reasoning to railway safety, but the distinct approaches have very different consequences for safety policy. This will be considered again below.

2.1.5 Multiplying VPF

The report mentions, but frustratingly doesn't explain, a special technique for approaching multiple fatality accidents. (p. 47) However it makes clear that there are limits to its application, and they do not seem to apply to railway safety. The only multiplier it recommends is in the case of risks of cancer, which are said to be especially feared. (p. 65)

2.2 Psychological Issues

There is a huge literature on the perception of risk. In the time available only a small proportion of the empirical literature could be reviewed. The main focus has been the work of Paul Slovic and associates, especially [Slovic, 2000a], which contains reprints of his major papers and [Flynn, Slovic and Kunreauther, 2001]. Slovic was selected as his work is referred to both by the Health and Safety Executive [R2P2] and in the Jones-Lee studies [Jones-Lee 2000a and Jones-Lee 2000b, also Marris, 1993]. It also seems well-researched, systematic, insightful and solidly argued. It is very helpful in the present context, even though the topics covered in this literature can provide only part of the analysis sought here. The work of Adams [Adams, 2001] has also been consulted, and while its relevance to these issues is less direct it has provided a very helpful perspective.

2.2.1 The perception of risk.

The psychology of the perception of risk seeks to explain variations in attitudes to risk, both between people and within any given person's individual attitudes. The question for the psychology of the perception of risk is what helps determine these variations.

It has become almost a dogma in the literature that the following factors can influence an individual's preparedness to accept risk [Slovic, Fisschhoff and Lichenstein, 2000a]:

- Familiarity: a new risk is typically perceived to be a greater threat than an old one.
- Voluntariness of exposure to hazard: people are more tolerant of risks where they can control their exposure to the hazard.
- Degree of control over risk: people are more tolerant of risks where they have some influence over the probability of the hazard occurring.
- Potential for catastrophe: people are less tolerant of risks that have the potential for causing catastrophe.
- Dread: certain risks induce greater 'dread' than others. For example people appear to be more afraid of dying of cancer than of other causes of death.
- How well known to science: people will in general be less tolerant of risk where the nature of hazard is not thought to be well-known to science.

There is also the obvious point that people will be tend to be more tolerant of risks to which their own personal or family exposure is low.

Finally, although not mentioned in the literature I have seen, the potential for a 'near miss' is often thought to be important in influencing attitudes to risk. This is well known to the gambling industry. A rail crash has high potential in this respect; many people will reflect that in other close

circumstances they would have been on the affected train, or know someone who might have been. This can make accidents especially real to people

Some of these points have clear application to railway safety – potential for catastrophe; control – whereas others may be irrelevant. All of these factors can influence perception of risk. However, note that all of these factors are determinants of individual perception of risk. If they were to exert an influence we would observe that people would be unreasonably afraid of railway travel. But we do not observe such a thing. Thus the psychological issues introduced so far cannot explain the extreme societal reaction, which is the main aspect of the puzzle.

2.2.2 Framing factors

Framing issues, which plague social science research, have a great potential to explain disparities in attitudes to risk. For there is evidence that people will give different answers to effectively the same question depending on how it is framed. Although this often occurs in the experimental situation, it also reflects 'real-life' biases. Of the various framing issues, three seem particularly pertinent.

- *Absolute versus relative judgements*. It is well known that if people are asked whether more public money should be spent on, say, higher education, most people will say yes. But if they are asked to order their priorities from a list including primary education, health care and law and order, then higher education might be a very low priority and actually may be thought not in need of further public money at all. In the case of railway safety we can, therefore, expect different responses in relativity studies to the responses we would get in free discussion concentrating on railways alone. But this is a perfectly general phenomenon and the same may happen for the discussion of any hazard. So, for example, there is no reason to think that road safety and rail safety should differ in this respect, but there is no evidence that the initial

conundrum affects road safety. Thus despite initial appearances this does not seem a fruitful line of investigation.

- *Quasi-elimination*. People are particularly concerned to eliminate risks completely, and so if a hazard can be entirely eliminated people have been found to have a preference to do so, even if the same resources could be used to reduce their overall exposure to risk – even exactly the same risks described a different way – to a greater degree. [Slovic,. Fischhoff and Lichenstein, 2000b] Suppose people face a hazard H. They are told that they can reduce the risk of this hazard by 50%. Let us suppose they are prepared to pay £x pounds for this risk reduction. Suppose, next, they are then told that there are two causes of the hazard, A and B, each responsible for half of the risk. They are then told that A can be completely eliminated. Typically, it is claimed, people will be prepared to pay more for the complete elimination of B than the reduction of the risk by 50%, even though it is the same action, described differently. This has very important applications to railway safety, as it is commonly thought that a whole class of accidents could be eliminated by the introduction of new signalling technology. The theory predicts that people would be far more likely to want money spent on this than they would be to see risks reduced to the same degree through other measures. It should be noted, however, that this does not seem to have been tested in the context of railway safety as the studies undertaken so far have not considered variations in the types of railway safety risks, except in terms of whether they can lead to multiple fatalities. Quasi-elimination plausibly predicts markedly distinct attitudes to different types of accident, in terms of how much money should be spent to avoid them.
- *Value of a loss*. Many experiments show that people pay more attention to losses than gains of similar magnitude, and will put a higher value on things they have than things they do not have. This is known as the endowment effect. Although to my knowledge this is untested, it is a reasonable conjecture that those who think that railway safety is in decline will be more inclined to want to see money spent on railway safety than those who think the system is stable or improving. Such effects are said to have been observed in studies looking at river pollution. [Slovic 2000c, p. 394]

In conclusion framing factors have a high potential to make a contribution to the analysis of the central problem. First of all, they typically show how it is not unusual for a given individual to have apparently inconsistent attitudes to risk. But more specifically to the present case, the framing factors of quasi-elimination and the value of a loss may play some role in explaining the extreme reaction.

2.2.3 Who should pay?

Again although little discussed in the literature there are obvious psychological and ethical questions about who should pay for increased railway safety: taxpayers; passengers or shareholders? How this may affect perceptions does not appear to have been studied, so far as I am aware, but there seems little doubt that people will be willing to see ever more spent to reduce a risk if they do not have to pay for it themselves. Arguably this may go some way to explaining the division in attitude that we note: people may give different answers to the questions: ‘how much would you pay?’ and ‘how much should be spent?’. Most important, though, is the question of how attitudes to risk would be influenced by the thought that safety is being sacrificed for the sake of profits. This will be discussed again in Part 4 and following.

2.2.4 Social amplification of risk

Although there are a number of factors that will explain why people may have two attitudes to risk, there is not enough here so far to explain why there has been such an increase in apparent societal concern for railway safety in a way in which there has not been for other risks (although air traffic control is apparently becoming of increasing concern). It appears that further factors must be incorporated to go further, and in particular the notion of the social amplification of risk may seem promising.

Although an elaborate theory of the social amplification of risk has been outlined, if not developed in detail, the salient point is simply that intense media scrutiny and reporting will make issues especially vivid to people, and this can lead to an intensifying of attitude. [Kasperson, 2000] A railway crash in the UK will attract enormous sustained and detailed coverage in the UK and, if major, will be reported in the world media. It will become a talking point and even technical details will become known. This may give rise to two further effects: first individuals may greatly over-estimate the risk, and adjust their behaviour accordingly (corresponding to increasing individual concern); and second, there may be increased societal concern to the effect that these sorts of things are a serious problem, even to those they don't affect personally. Although social amplification can, in theory, be the result of a number of mechanisms, it seems clear that in the case of the railway safety the media are largely responsible for making the issue salient to so many people.

2.2.5 Signal value

It is likely that any railway accident involving the death of more than one passenger will receive the media attention which will amplify risk at the individual level. However, this effect may well dampen down quite quickly, although in certain circumstances the initial amplification may be reinforced by further features. In particular certain accidents have 'signal value'; that an accident shows that a new type of risk – hitherto unknown or unappreciated by the industry or by the public – is possible. [Slovic, Fisschhoff, and Lichenstein, 2000a]. When combined with extensive media coverage we have a very potent mix. The accidents at Southall, Ladbroke Grove, Hatfield and Potters Bar, among others, are arguably all of this nature, at least in part. What it means is that people come to believe that past statistics are no guide to future safety. It will also lead to the thought that those who manage the system are not to be trusted. Either they were not even aware of the risks involved, or they knew about them but 'didn't care'. This calls the whole system of risk management into question, which is exactly what we have seen. Confidence falls even lower when

a second incident of the 'signalled' type occurs, as was the case when the Ladbroke Grove accident took place two years after Southall.

2.2.6 Systems failure

The cases mentioned in the last section all have something further in common; they all resulted from what are publicly regarded as 'system failures'. We need to distinguish between two types of cases; first those caused by factors perceived to be external to the system, e.g. by vandalism, trespass (knowing or accidental) or by factors that are an inevitable part of the risks of life, such as trips and falls; and second those which appear to be attributable to some sort of systems failure, or design fault. There are two aspects of system failure that are highly pertinent; first the idea in the public mind that they could be completely eliminated by known technology, or known techniques of maintenance; and second the idea that it is a failure of duty within the industry that they are not eliminated. This is central to the analysis that follows.

It is, at this point, worth introducing a further distinction between two types of systems failure. First there are cases where a system is in place but it fails through avoidable error: either operator failure; poor management; or design fault. This, then, is *failure of a system*. Second there are cases where the failure is that a system using known technology could have been introduced, but, for whatever reason, has not been so. This is a *failure to introduce a system*. This distinction is somewhat complicated by the fact that the call for the introduction of a new system is often the consequence of the failure of an existing system. However there are clear cases where it is thought that the proper application of, or small refinement to, an existing system would have prevented an accident, and there is no call for thorough-going change. This distinction will often correlate with the difference in the cost and disruption of making the called-for changes.

It is understood that the distinction between those risks thought to be under the industry's control, and those which are not, or are less so, is well-known, and an attempt is made to take this into

account [Railway Safety, 2002d]. However the importance of this factor may have been seriously under-estimated to date, and we will return to it below.

It is worth noting that in various documents [eg Railtrack 2001] a great deal is made of the distinction between single-fatality accidents and those which have the potential for causing multiple-fatalities, also referred to as catastrophic risks. However it is arguable that this distinction is far less significant than that between risks caused by systems failures - similar to the idea of those under the industry's control - and other risks. On the analysis presented here the main significance of multiple-fatality accidents is that they will attract immense media attraction. However the further consequences of such attention will be dependent on other factors.

2.2.7 Stigma

Social amplification can lead to stigma where a product or technology is treated as to be avoided to a degree out of all proportion to its risks. Nuclear power is the most often cited example, although 'chemicals' are often treated in a similar ways. Particular products can become stigmatized; it is said that the episode in which Tylenol was tampered with, leading to seven deaths, also gave rise to 125,000 news stories and cost Johnson and Johnson around one billion dollars. [Mitchell, 2001] The railways have from time to time appeared to be moving in this direction, but at present there is little evidence that people have an irrational fear of railway travel.

2.2.8 Conclusions

The industry can count itself as extremely unfortunate to have had a spate of accidents which combine perceived systems failure, signal value, quasi-elimination and value of a loss , all of which have taken place under the intense social amplifying mechanism of the media. Yet according to standard theory most of these factors should lead to heightened individual concern for

railway risk. However there is little evidence of increased sustained concern at the level of individual risk. Rather there appears to be a level of societal concern which is not yet fully analysed. It is to this we now must turn.

Part 3. Analysis

Psychological factors alone have not been found sufficient to understand the central problem; that people can believe that the railways are very safe, yet insist that vast sums should be spent on safety improvements. In the following analysis it is suggested that this cannot be understood without attributing particular moral assumptions to those members of the public who call for substantial further investment in safety. Existing conceptual frameworks of risk analysis typically do not make room for such factors, and this is why they are unable to explain the phenomenon. Once these moral assumptions are isolated they can be examined, which will both deepen the analysis and point towards a possible resolution.

3.1 The Ethics of Risk: Consequentialism and Absolutism

The central dilemma can be put in terms of a conflict between ethical approaches to risk. A consequentialist approach evaluates actions or decisions purely on the basis of the consequences to which they lead. An absolutist approach, on the other hand, rules that some actions or outcomes are required, or, conversely, are unacceptable, independently of the further consequences to which they lead.

A consequentialist theory can be divided into two parts. First it has a theory of the good: an account of what is to be considered as valuable. Second it has a theory of the right: what we should do with respect to the good. Utilitarianism is the best known consequentialist theory. The good, according to classical utilitarianism is pleasure and the absence of pain, and the right action is that which maximizes the balance of pleasure over pain. It is often assumed that consequentialism must be a maximizing doctrine. However, this is not strictly correct. Evaluating actions in terms of their consequences is not the same as supposing that the right action is the one that maximizes the expected balance of good consequences. Other functions, which may, for example, include a distributional element, may also be possible.

It is essential to consequentialism, at least in most forms, that the good can be quantified, for without quantification the notion of maximizing makes no sense. Even those forms of consequentialism that do not appeal to the idea of maximization will still need some form of ordering. This is a major difficulty for philosophical utilitarianism, requiring the solution of the problem of 'inter-personal comparisons of utility'. Consequentialist approaches to public policy decision making side-step this problem by translating everything to be taken into account into cash terms, and then pursuing whichever option provides the greatest expected profit or smallest costs so defined. It should be noted that such 'utilitarianism of money' is not identical with classical utilitarianism, and may not yield the same results. For example, it is normally assumed that people obtain diminishing marginal utility from money, and so utilitarianism would generally favour more equal distributions than cost-benefit analysis, in that equal divisions will typically yield a greater total of utility than unequal divisions of the same total sum of money. Hence standard cost-benefit analysis is only an approximate form of utilitarianism.

Applying consequentialist reasoning to railway safety could be attempted in many ways, of which standard cost-benefit analysis is only one. An alternative might be to try to translate everything into degrees of preference satisfaction, and to maximize this value, or to try to define a further notion such as 'intrinsic reward' [Gibbard, 1986]. However this has obvious difficulties and cost-benefit analysis has the merit of providing a more straightforward approach. To apply it to safety issues requires one to think in terms of 'value for money': we need to determine how much safety, and how much money, invested in which ways, provides the most favourable outcome. In effect it balances a desire for safety against other factors; essentially the opportunity cost of spending money on particular safety improvements - what else could have been done with that money.

Within this framework it makes perfect sense to ask how much it is worth to save a (statistical) life, and to rule out some proposals as out of proportion, as being too expensive. In effect this is to make the judgement that there are alternative, more efficient ways of turning money into beneficial outcomes, including more efficient savings of lives.

Clearly it is central to risk cost-benefit analysis of this nature that some sort of financial valuation is put on the value of life, so that it can be traded or balanced against other factors. This is sometimes used as an objection to consequentialism. However, the issues are conceptually distinct. It is possible to object to the idea of trading off a life against other values yet remain within a consequentialist framework. For example, if a theorist was to claim that the sole good was the preservation of life, and we should do all we can to maximize the preservation of life, this would still be a consequentialist theory, albeit a very implausible one as it would always require the sacrifice of the quality of life in favour of its preservation. Applying such a view to railway safety would appear, in the first instance, to suggest that we should do everything in our power to reduce railway deaths to zero. However, this is probably not a view that anyone holds. If, for example, we could come closer to this result by operating with a maximum speed limit of 10 mph on all routes probably no one would say that this is a price worth paying.

Absolutist theories are harder to define, although they have in common the rejection of pure consequentialism. At the least they say that actions and decisions cannot always be justified purely in terms of their consequences. One standard way of characterizing absolutist theories (or deontological theories, as they are known in the philosophical literature) is that they typically define the right independently of the good. To explain, while consequentialist theories define right action as maximizing the good, absolutist theories define right action in some other way: perhaps as following a set of rules, given by god or by reason or tradition, or even by ‘common sense’.

Accordingly absolutism will tend to focus on particular types of actions, and rule that these are impermissible, at least in all normal circumstances. So, for example, most people will accept that deliberately killing an innocent person is not acceptable, even if we know that greater good will come of it (for example, we would not kill an individual in order to save four other lives through organ transplants).

This example illustrates the point that it is unlikely that anyone can be represented as being a pure consequentialist or pure absolutist. Rather we are all used to employing both consequentialism and absolutism in our ordinary ethical thought; sometimes even being conflicted between absolutist and consequentialist elements in a single example (for example, over the question of when it can right to break a promise).

From the absolutist perspective the moral assessment of an action will depend crucially on the proper description of that action. An accidental death and a murder both have the same direct consequence: someone dies. From an absolutist perspective they have a quite different moral character.

So absolutism pays attention not merely to outcomes, but to how outcomes are produced. In the case of railway safety, absolutism will not only want to focus on the number of deaths, but the manner in which they were caused. It is quite possible, on this view, that absolutism will recommend putting resources into reducing one type of accident, even though using those resources some other way would provide a greater risk reduction. From a consequentialist perspective this seems quite irrational, but it is a normal part of our ethical thought. For example, the resources currently devoted to preventing murder might save more innocent lives if they were put into road safety or the health service. Yet it seems perfectly reasonable to think that reducing murder ought to be our priority (and not only because we worry about the longer-term consequences of murder getting out of hand). This, of course, presupposes that we can grade types of actions according to their moral character, where that character is

not (wholly) determined by their consequences. It also presupposes that we have a special responsibility to root out behaviour believed to be (seriously) immoral.

From the existing studies it seems that some, perhaps many, people flit easily between consequentialist and absolutist attitudes in thinking about railway safety. On the one hand people can be brought to take essentially a consequentialist attitude to railway safety, thinking that it needs to be balanced against other factors, and that further improvements are a low priority. Yet on the other hand, the same people may suppose that the behaviour of the industry verges on the criminal if it doesn't take very expensive measures. To anticipate what is to follow, I will suggest that people are prepared to take a consequentialist attitude when thinking about their personal exposure to risk, and which of the risks they face in their own lives they are most concerned about. However, they will take more of an absolutist stance when thinking about the types of risk that should exist in their society. This corresponds to the distinction between individual and societal concern, and again to a distinction between consumer and citizen.

3.2 Individual And Societal Concern Revisited

To provide an analysis of the conundrum we need to combine aspects of the psychological literature with a better moral understanding of the notion of societal concern. The HSE account of societal concern is, as suggested in Part 2, not entirely clear. Indeed the conceptual framework of risk analysis stemming from the HSE may be seriously incomplete, in the sense of making too much of some aspects of risk and not enough of others. To bring this out, consider one example the HSE itself uses. The following is quoted from R2P2.

> The actual risk of death per annum for the public from work activities is usually very much lower than the figure of 1 in 10 000. For example during the period 1994/5 to 1998/9 the annual risk of death to the public from the use of gas (fire explosion or carbon monoxide

> poisoning) averaged over the entire population of Great Britain was 1 in 510 000 - in other words below the limit of what is regarded as broadly acceptable. Gas incidents, however, continue to give rise to societal concern, particularly where the incidents occur because unscrupulous landlords seek to avoid the cost of simple safety checks on their gas heating systems and so put those who rent the accommodation (often young people) at greater risk. In effect such societal concerns override averaged numerical considerations. HSE has responded by firm enforcement action where appropriate, and by targeted publicity emphasising the importance of annual safety checks on gas appliances. (p. 44)

There are several points worth making here.

1. Carbon monoxide poisoning in rental accommodation is given as an example of a risk arising from a work activity, but this seems highly misleading.

2. The notion of societal concern is used here, yet it does not obviously correspond to HSE's own definition. Some attempt seems to be made to suggest that societal concern is engaged because 'young people' (thus a vulnerable group) are put at greater risk but this does not seem to get to the heart of the issue.

3. Rather, as the discussion itself makes clear, the key point is that the risks are imposed by 'unscrupulous landlords'. But without acknowledging it this seems to introduce a new and vitally important element: **that of who or what is exposing people to the risk, and with what moral justification.** It seems that it is because we are appalled by certain landlords' irresponsible behaviour in avoiding 'simple safety checks' that societal concerns are engaged. Here, then, we see ethical issues, of a particular, absolutist flavour, appearing as central to the analysis of attitudes of risk.

To capture this a more developed set of distinctions is necessary to replace, or at least enrich, the existing distinction between individual and societal concern. The analysis should be framed by two questions: who is concerned and what are they concerned about?

The likely answers to the first question are: 'individuals'; 'society' and 'government and its agencies'. It is an interesting philosophical question whether society can have concerns that are not reducible to individual concerns, but it is not one we need to address here, as we can work with a framework in which only individuals have direct concerns. (Governments will have the indirect concern of how well they are perceived as responding to individual direct concern.)

Concentrating on the concerns of individuals, we can now ask the second question: what are they concerned about? In the first instance they may be concerned about the harm that may be done to people purely in virtue of the risks they face. Here we can make a distinction between a number of cases. Personal concern is a matter of concern about risks for oneself and perhaps one's household. Preferential concern is concern for risks to wider family, friends and acquaintances. Extended group concern is concern for risks to a group of which one is a member. Altruistic group concern is concern for risks to a group of which one is not a member. Altruistic concern is concern for risk to people generally.

Yet there is a further dimension to consider, where the concern is not focused on the harm that may be suffered, but on the fact that certain hazards exist and risks imposed in the way they are. As we saw in the gas poisoning case, concern is engaged not so much because a very small number of young people will die, but because these risks are being imposed irresponsibly, by 'unscrupulous landlords'. If risks are perceived to be the result of reckless, negligent, incompetent or selfish behaviour this moral issue can become a matter of great concern, **independently of any concern for the actual harm likely to be done**. It is at this point, it seems, that we can talk of a form of societal concern which is not merely altruistic concern. It looks not so much to those who may

suffer, but those who are imposing risks on others. It is an essentially moralized notion in that it is premised on a morally critical attitude to those imposing the risk.

Here we can now see that initial HSE conceptual framework is not sufficient to capture all the salient features relevant to risk perception (and this also appears to be a failing in the psychological literature). In addition to the concepts of hazard and risk we need, in at least some cases, to take a third element into account: the risk imposer, or the process by which the risk is created. Just as a particular hazard may be thought to be unacceptable, independently of any consideration about the level of risk to which it gives rise, a certain risk imposition may be thought to be unacceptable, independently (to a degree) both of the hazard and the level of risk to which it gives rise. This, we saw, gives rise to a form of societal concern, in which people are concerned about what their society contains or has come to, which is distinct from concern about threats of harm to people.

Where there is such apparently unjustifiable imposition of risk, individuals and groups may then put pressure on government or other institutions to do more to control it, and these institutions will be thought to be at serious fault if they do not act decisively. Any resulting accidents will be treated as scandals for those regulating risks as well as those imposing them. This, then, causes government concern, which is a distinct form of societal concern, of a sort closer to that defined by HSE.

3.3 The Ethical Limits of Risk Cost-Benefit Analysis

Standard risk cost-benefit analysis does not take account of societal concern where such concern is based on the manner in which hazards are produced, as distinct from the nature and risk of the hazard. To understand this we need to look at the ethical foundations of cost-benefit analysis as applied to risk. There are three stages to the analysis in this section. The first responds to a common suggestion that risk cost-benefit analysis is not based on any ethical assumptions at all.

The second rebuts this suggestion, attributing a consequentialist foundations to risk cost-benefit analysis. The third explains why this approach leaves us unable to understand the central dilemma.

3.3.1 Does Risk Cost-Benefit Analysis Make Ethical Assumptions?

The case that risk cost-benefit analysis is not based on any ethical view stems from the claim that the point of risk regulation is to address a certain type of market failure. To understand this we need to ask what is, perhaps, the most fundamental question regarding the safety of any activity: why should safety improvements ever be made? Of course in the current climate there is a statutory duty to address the safety of certain commercial activities, but let us consider a hypothetical situation in which there is no safety legislation. What would be the problem? Presumably it would be said that people will be subjected to risks that they are in no position to assess, and might wish to avoid if they knew about them. So let us imagine that everyone knows exactly the hazards and risks each activity involves. Let us further assume that they meet certain standards of rational consistency in their preferences about risk. In such a world people can take informed decisions about what risks to face, and risk would become one factor people will take into account in deciding whether to use a product or service. Would there be any further reason for regulating? Perhaps there would, because of the problem of externalities: people facing risks that are the consequences of others' actions and decisions. But leaving this to one side, it could be argued that if people knew all the risks they faced, they could simply make their own decisions, and the economy would be self-regulating in this respect. If an activity was seen as too risky, custom would fall, and thus if the activity is to continue, safety improvements would be needed to bring customers in (or the price of the activity would have to fall). Alternatively if the activity is safer than it needs to be, this may mean that the price is higher than it needs to be and customers would depart in favour of firms who manage the risk/cost trade-off in a different way.

Consequently in the first instance the regulation of safety can indeed be seen as a response to a certain sort of market failure: that consumers do not have good information about the risks they face. For if they did they would take for themselves the decisions that the Health and Safety Executive attempts to take on their behalf. We can see the situation as analogous to product quality standards, and other forms of consumer protection, which are responsive to the fact that consumers are likely to be ignorant of many salient factors. Here, again, it is arguable that if consumers were omniscient, regulation would be redundant, for people could make their own choices. Of course the question of externalities – risks to third parties - complicates the analysis, but these can be handled also as market failures, to be solved by compensation.

This impression that risk cost-benefit analysis is simply a means of responding to market failure is reinforced by considering that the approach regulates risk on the basis of what people would be prepared to pay to avoid risk. But if there was complete knowledge, the market would be the test of what people are prepared to pay. A profit-maximizing company would automatically seek to find the optimum price/safety trade-off, in the awareness that setting the trade off at the wrong point will lead to reduced profit. There would be no gap between what people would, on average, pay to avoid risk, and what they do actually pay in a perfect knowledge market. Thus, the 'risk market' would equilibrate for itself; converging on a true VPF at profit-maximising price. (This, importantly, requires some modification in the case of a monopoly supplier, but this point will not be developed further here.)

3.3.2 The Ethical Foundations of Risk Cost-Benefit Analysis

Even if it is true that the point of RCBA is to address market failure, it does not follow that it is not based on any ethical assumptions. For one has to ask why it is thought important that efficient market outcomes should be pursued. Efficient market outcomes are, broadly, those where people in the aggregate get more of what they want than they could do with any other

use of the given resources. Why does this matter? There seems no alternative but to admit that this involves an appeal to the idea that preference satisfaction is good in itself, and the more of it the better. This is, of course, the moral theory of preference utilitarianism, which can thus be seen as the ethical foundation of standard risk cost-benefit analysis. Further ethical assumptions are brought in by the HSE modifications to the calculations in the name of equity, but at bottom risk cost-benefit analysis is clearly a form of consequentialism.

However a quite different ethical framework has also been claimed; RCBA analysis has been said to embody a principle of hypothetical consent. [Leonard and Zechkauser, 1986] That is, it produces the effects which people would agree to had they been in full possession of the facts. This makes it appear ethically very robust indeed, drawing on two powerful ethical traditions of utilitarianism and the theory of the social contract, which often appear to conflict. That is, the main objection to utilitarianism is that it victimises individuals, whereas contractualism requires the agreement of all.

3.3.3 Factoring in Societal Concern

Nevertheless the sense in which CBA models hypothetical consent is very restricted. We have to ask what it is that people are supposed to be giving their hypothetical consent to. RCBA models an implied willingness to bear certain risks. However, and this is the key point, **even if an individual is willing, in the circumstances, to bear a risk (i.e. not to pay to avoid it), it does not follow that this individual thinks it acceptable that the risk exists in their society**.

To illustrate the problem, consider the issue of whether an individual is prepared to take the risk of walking home from the station late at night, believing that there is a 1 in 250,000

chance of being murdered, when the alternative is to take a taxi at the fare of £5, which, let us say, reduces the chance of death on the way home to 1 in 10,000,000. Now, if someone decides to walk, what can we infer? It seems uncontroversial to say that on this particular occasion that person was prepared to bear a particular risk of being murdered for the sake of not spending £5. However, can we go further? Any further conclusion goes beyond the immediate evidence and thus must be based on some assumption or other, or, indeed, on a cluster of assumptions. Could we argue that this decision shows something about the limits this person thinks society should pay to try to reduce the murder rate? If we try to argue this it is based on the assumption that the amount someone is prepared to pay on a single occasion (or perhaps, on average) to avoid a risk can be used to determine how much they think society should pay to reduce it.

However this inference seems to me to be a mistake, for the following reason. When you are considering whether or not personally to bear a risk, the important issue concerns the nature of the hazards and their likelihood. Rational decision making does not need to investigate the processes by which these outcomes are produced (except to the extent that this will help you formulate a strategy to avoid it.). However, when you are considering whether a risk should exist in your society you will also want to know the details of the process which produces that risk. In particular the moral character of that process will influence your decision. Socially, we are typically much more concerned about eliminating murders than accidental deaths, although from a given individual's point of view a death is a death, and it is perfectly rational to concentrate on such outcomes to the exclusion of what may have caused them. Of course some ways of dying are more feared than others but this is a separate issue. It makes perfect sense to believe that we should, as a society, do more (or conceivably less) to prevent a death by arson than a death by accidental fire even if both are feared equally by individuals.

My claim, then, is **there no direct link between the way people deal with hazards in their own life and their broader social tolerance of risk and hazard**. The first is a matter of making the best of choices within a framework, the second of whether elements of that framework should exist.

We have already distinguished between individual and societal concerns. Individual concerns are captured by RCBA's averaged figure of people's preparedness to pay to avoid a risk. This corresponds to the consequentialist element in a given individual's ethical thought. Indeed risk cost-benefit analysis, and its underlying consequentialism, is a perfect mechanism for modeling individual attitudes to risk. Utilitarianism is sometimes said to model social decision making as an extension of individual rational decision making, and this is exactly what we see here. By exploring individual willingness to pay to avoid risks we construct a type of social average hypothetical price. We then attempt to regulate so that individuals receive the trade-off they would have purchased (on average) had they been in full possession of the facts. Nevertheless we can easily see the limits of this approach. Societal concerns are not captured as they are based not purely on outcomes but, at least in part, on the processes by which those outcomes are produced. Where there is a belief that the processes are morally wrong there is a belief in culpability, which is the basis of societal concern. This corresponds to the absolutist element in ethical thought.

Another way of putting this is that there is a distinction between how we think as consumers and how we think as citizens. As consumers we make our choices from among the array of options we face. As citizens we are much more concerned about the profile of that array even when we have no personal interest in some of the options: does it reflect how a good society should operate? Accordingly we have three distinctions which appear to be related to each other: on one side consequentialism, individual concern and thinking like a consumer, on the other absolutism, societal concern and thinking like a citizen.

3.4 The Ethics of Societal Concern

Two questions arise immediately from this analysis. First, what types of processes of producing risks will attract absolutist societal concern? Second, how should we take societal concern into account in decision making? After all we cannot, as a society, put all our resources into, say, reducing murder.

Neither of these problems has an easy solution. There does not seem to be a known general approach to the question of how to take societal concern of this nature into account, other than the use of a multiplier, which I argue in Part 4 is not satisfactory. While in Part 4 I outline an approach that could be taken regarding railway safety, no claims are made about how far this may generalise. I will not attempt to say any more about this here, as it is beyond the remit of the current report. However we should note that it is a matter of great significance, and will need attention in future work.

The problem of which processes should attract societal concern appears at first to be a relatively straightforward matter of dividing processes into those which are morally suspect and those which are morally acceptable. One would hope that in a rational moral system it should be possible to set out such categories in a clear and principled way. However, human morality may not be so simple. There may be no such clear principles available. In a moral study of risk the philosopher Annette Baier points out that human beings impose risks and harms on each other in many ways, and many of these are not even noticed. Her claim is that ‘Morality is the culturally acquired art of selecting which harms to worry about, where the worry takes the form of bad conscience or resentment’. [Baier, 1986, p. 49] To put this another way, not all harms are considered to be wrongs.

On Baier's view historical and cultural factors will play a significant part in determining which processes of producing risk and harm we find unacceptable. To illustrate her point, in contemporary economic life we treat the financial losses, and thus harms, caused by economic competition as simply part of normal life, even though it can lead to bankruptcy and suicide. So this is a way of harming people that we have chosen not to notice, whereas similar harms caused by theft are treated with great seriousness. Actions can change status; consider the contrasting fortunes of marital rape and consensual homosexuality. Consider, too, the moral acceptability of certain self-harms, such as suicide, and the conditions under which this is considered morally wrong.

If this is right there may be no systematic answer to the question of which harms particularly raise societal concern. Baier's view is not the only possible one, of course, and theorists may rise to the challenge of attempting to lay out conditions by which we can classify actions into moral categories. However, this is fraught with difficulties, and I think it is safe to say that there is no generally accepted answer, although an enormous amount of effort has been spent on attempting to delineate the categories of common sense morality in a principled way. Note, though, that this literature is primarily directed to understanding the principles that underlie agreed moral intuitions. Finding such principles and applying them to problem cases (e.g. whether euthanasia is justified) very rarely leads to results which gain consensus.

Nevertheless the task of sorting processes into moral categories is not entirely arbitrary, and the standards of common sense morality and consequentialism will converge on some significant categories. For example, we will think culpable those who deliberately impose easily avoidable risks on others. We are likely to disapprove of those who knowingly impose risks on others as a way of 'dumping' their costs. Common sense morality and consequentialism will agree that many risk impositions of this nature are unacceptable.

There are also cases which are very clearly unacceptable from the point of view of common sense morality, although the consequentialist analysis may be less certain. Deliberate or malicious harming of others is generally regarded as seriously wrong (with some exceptions for judicial punishment) irrespective of the consequences. These are areas where, as noted, standard risk cost benefit analysis, because it is focused on outcomes rather than their causes, may find it very hard to explain why we care so much about such outcomes caused one way rather than another, although common sense morality is in no doubt.

The most difficult category, however seems to be that of imposition of risk and harm which appears negligent or reckless. The difficulty is that not all unthinking or even knowing imposition of harm, or risk, is commonly regarded as culpable. The economic competition example illustrates this. Consequentialism may be thought to provide the means to make this distinction (economic competition is allowed because of its long-term beneficial consequences), but the case of railway safety shows this does not generalize to every example. For the whole of our problem is that it appears that some significant part of the public believe that the industry is culpably negligent if it does not adopt certain measures, even though such measures cannot be justified in consequentialist terms. A calculation of consequences would justify a particular level of safety, but it appears that these members of the public think that the industry is culpably negligent if it fails to adopt certain very much more costly measures. So either the consequentialist calculation has gone seriously wrong, or consequentialism cannot explain public belief in culpability.

It is within this area of imputed recklessness and negligence that belief in culpability appears most likely to be most culturally dependent, even though such dependence will also apply to the classification of some deliberate harms. What we choose to notice and care about may, in some cases at least, be a matter of selection. Something taken for granted in one society or generation (physical punishments, for example) might appear quite unacceptable in another. Furthermore, it may even be that measures taken by one type of organisation – a private

company – may appear culpable when the identical action taken by a non-profit-making organisation may appear perfectly reasonable. At the least we cannot rule this out.

One promising line of thought in this respect is to consider the cost of preventing the hazard in proportion to the severity of the hazard, leaving aside the risk. To explain, consider again the carbon monoxide case. Here a standard cost-benefit analysis may conclude that there is no reason to insist that landlords make gas safety checks. For purposes of illustration, suppose the chances of anyone dying from such poisoning in a given period is one in a million, in that one million people are exposed to this risk in rented accommodation and just one will die. If a gas safety check costs £50, then insisting on a million such checks gives a VPF of £50m, which is way above the standard figure of around £1m. However it seems that in this case we treat the probability of death as virtually immaterial to the analysis, looking only to the hazard and the cost of preventing it, as it falls on each particular individual imposer of risk, rather than in total. Although this is hard to state in any precise way, it appears that we believe that where the costs for an individual risk imposer of taking steps to prevent the hazard are small, and the hazard is severe, involving death or serious injury, then we believe that a failure to take these steps is culpably negligent, especially when imposed as part of a commercial activity. This we could call the 'proportionality principle'. It clearly has implications for railway safety, which will be explored in Section 3.6. below.

3.5 Social Amplification: Risk and Culpability

We are now close to being able to provide a complete analysis of the initial problem, but to do so we first need to draw in the idea of social amplification. In the normal course of events people will come to be informed about the risks they face, and who or what is imposing these risks on them, largely through the media. Thus the risks dealt with by the media will take on special significance. Immediately, then, this will bring into play mechanisms of social amplification. We should note, though, that there seem to be two distinct ways in which this can operate. In the standard case,

which is the one presented and studied in the literature, individuals will experience heightened concern about the risks to themselves and adjust their behaviour accordingly (e.g. move away from a district where a nuclear power station has been sited, or stop buying certain foods). Societal concern may also increase, reflecting increased anxiety about whether greatly feared risks are being taken seriously by those imposing and regulating them. However, in the case of railway safety a distinct mechanism of amplification seems to be at work. That is, it appears that although personal and altruistic concern rises immediately after a prominent accident it very soon dampens down again. People rarely adjust their lives to avoid the danger of a train journey. Societal concern, however, rises enormously and stays high. Societal concern seems, then, to be floating free from individual concern. **What we observe here is not so much social amplification of risk, but social amplification of moral culpability**. We could say that the public have come to think that the state of railway safety is a national scandal; it is this reaction which leads some people to think that vast expense should be made to prevent certain types of accident. However if it is right that this is a societal concern that is not based upon individual concern about risk, as distinct from the irresponsibility with which the risk is imposed, it will not be picked up in studies designed to measure individual concern.

3.6 Culpable Processes and Systems Failure

From the previous analysis it appears that at the root of the issue is the public perception that the railways are failing in their safety duties. Yet this perception is clearly problematic. To the extent that it follows the ALARP principle in practice, the railway industry does appear to be carrying out its duties, as set out by HSE. So why does the public believe that it is neglecting its duties?

First of all, very few people have any idea at all that the industry takes any systematic approach to safety. There is, then, some possibility that the sense of neglect would disappear if it became public knowledge that the industry uses the type of RCBA that it does. However, we have no evidence that the public would accept the ALARP principle even if they knew about it. Rather, it

seems likely that the public have a more complex response, and will tend to differentiate types of accidents and the appropriate response. In particular, it is a reasonable conjecture that for 'ordinary' risks the ALARP principle may carry some conviction, but for risks relating to systems failures this would seem quite unacceptable.

However, the application of the 'proportionality principle' set about in Section 3.4 may be helpful here, for it has consequences for the forms of system failures that can reasonably be represented as culpably negligent. The basic idea is that if a hazard involves risk or death of serious injury, and this risk can be reduced or even eliminated, relatively cheaply, then such steps should be taken, even if the risk is already so low that cost benefit analysis does not recommend taking such steps. If such risks are known about, yet not taken, then it is arguable that the risk imposer is culpably negligent. Recall also a distinction has been made between two types of system failure: failure of a system, and failure to introduce a new system. Typically the preventing a failure of an existing system is likely to be relatively cheap, or at least perceived as such, whereas the introduction of a new system is likely to be relatively expensive. Thus accusations of culpable negligence seem more likely to be firmly based in the case of the failure of an existing system. (However we need to bear in mind that the proportionality principle appears most convincing where the individual costs of taking precautions are widely spread, and so no one bears the aggregate cost of many individual risk reductions. In cases where a single agent bears all of the small individual costs this may aggregate to a very large total. Here – which is exactly the case of the railway industry – our moral intuitions may be far less clear).

It does not follow, however, that this exhausts the category of culpable negligence, for there may be other tests. For example, there may be a 'benchmark of good practice' test, where one is perceived as failing if one does not adopt systems adopted by comparable organisations.

And there may be other tests too. There is some reason, then, for thinking that a further study of the moral and legal ideas of culpable negligence could shed further light on these issues (this is discussed further in Part 4).

Indeed it seems likely that many members of the public believe that the industry is culpably negligent in respect of both types of system failure. In the case of failure of a system it might be supposed to fail the proportionality principle. In the case of failure to introduce a new system it might be suppsoed to fail the 'benchmark' test. In this latter case, it will be assumed that 'state-of-the-art' is the appropriate approach to risk control in the case of railway safety, at least for those risks perceived to be under the industry's control. Where the industry fails to meet this standard it is regarded as culpably neglecting its moral duty. However this aspect of public belief has been little studied and so exactly why this is believed (if it is) is largely a matter of speculation. The best evidence of public attitudes we have is the qualitative discussion within the Jones-Lee follow-up study [Jones-Lee, 2000b]. At least some people within this study believe that not only is such technology available but that it is in common use elsewhere, and thus the UK suffers in this benchmark for acceptable practice. In addition, the framing bias of quasi-elimination (the preference for eliminating entirely one source of risk) will affect societal concern as well as individual concern; people will be keen to see one hazard eliminated, even if this is an inefficient way of reducing total exposure to risk.

Furthermore many people believe that things are being allowed to get worse, which adds to the idea of culpability, while also bringing into play the further framing issue of 'value of a loss'. When this is tied in with the idea that shareholders are profiting at the expense of safety, the behaviour of the industry may appear quite shocking to some people.

It is striking that the general idea of weighing the costs of railway safety improvements against its benefits seems not even to occur to most people. One possible reason for this is another shortcoming of the HSE-style approach as applied to railway safety. That is, that this approach is

designed to regulate risks arising from work. Yet for the train passenger the process is more like buying a service than being exposed to a risk from work. So a better paradigm may be decision criteria used in product safety, or the regulation of food and medicine where far higher standards of regulation may well be in use, although this has not been examined for the purposes of this study.

Consider the issue of road safety. This falls into at least two areas: safety of the infrastructure (junction design etc) and safety of vehicles. Although something like the ALARP principle may apply to the former, it seems unlikely to apply to the latter. For example it is a fair assumption that safety recalls are not always justified by the ALARP principle based on VPF, or that safety modifications in new models can be justified in this way. Probably a vast amount is spent on road safety, or at least the apparent safety of those driving the car [Adams, 2001], even though much of it will not appear in public figures as it is private spending.

In sum, media attention has led people to believe that the railway safety is in crisis; not because of the degree of risk, but because of the morally irresponsible way in which the industry is thought to be imposing these risks; essentially there is an allegation of culpable negligence, based on both the proportionality principle, and the benchmark test. This applies in particular to risks of system failures, which, it is supposed by the public, could be eliminated by better management and the introduction of new technology. It may also be that it is thought that we lag behind comparable nations, and that this is the result of neglect, from which certain people profit, and from which it is necessary to ‘catch up’. The combination of quasi-elimination, value of a loss, international comparisons, ‘profits before safety’, and lack of acceptance of a CBA approach has, it is conjectured, through media amplification made the public come to believe that the railways are failing in their duties by recklessly or cynically imposing risks on the travelling public. At root, then, is the public perception of the moral failings of the railway industry. If the public came to believe that the industry were following the duties laid down by the HSE this may well lead to a spread of suspicion to the HSE, rather than to reverse the sense of neglect.

Earlier we noted the idea of social amplification of culpability. However, we can also consider whether there are cases where culpability is not merely amplified, but first constructed and then amplified. Somehow behaviour that was once accepted as at worse regrettable becomes unacceptable, and this then becomes a matter of grave public concern. This may be done through a confusion of cases where the allegation of culpable negligence is more plausible with those where it is less plausible.

In some cases of this type we may feel that what has happened is that we have come to object to something that we should have objected to all along: culpability is discovered, rather than constructed. In other cases, the claim of culpability is harder to defend, and may even look absurd when subjected to rational scrutiny; this is the realm of 'witch-hunts' and 'media-storm'. Therefore even if there is a general belief in culpability it does not mean that such a belief is justified. It may be based on false or misleading information, on ethical assumptions that are contestable or unjustified, or have no rational basis at all, being the result of journalistic or political manipulation. In other cases, the position may be arguable, with reasons on both sides. Unfortunately there will always be difficult and contested cases. However, the analysis presented here will assist reasoning about particular cases, and the further philosophical study proposed below may help further.

Nevertheless, we should note that even where the claim of culpability cannot be justified it can still be very damaging, and thus needs to be addressed. It may not be enough to be convinced that common social attitudes are unfounded. In the case of railway safety we do not know enough about public beliefs to know why it is that the charge of culpable negligence is made, although certain causes have been conjectured here. The point of the proposed sensitivity study (see Part 4) is to gain a better understanding of the causes of such a belief. It may turn out that the public belief in culpable negligence is based on a single false factual assumption, which can then be challenged. More likely it is based on a combination of factors, which may weigh differently with different people. Such an investigation will provide

a preliminary to a debate about how well founded this attitude is. Where it is ill-founded, it should be countered by good public relations initiatives. Where well-founded – i.e. where it can be agreed that certain processes are unacceptable, at least in part independently of their consequences – further consideration needs to be given to how it should be taken into account in long-term planning.

3.7 Implications of the Analysis

If this analysis is correct our questions become:

- what can be done to address the sense of culpable negligence?
- how should such societal concern be reflected in railway safety decision making?

The latter poses a dilemma for the industry. On the one hand replacing the RCBA approach with a state-of-the-art approach will be incredibly expensive. On the other hand, as things stand, if the industry fails to do this then it will come under intense pressure and scrutiny every time a 'preventable' multi-fatality accident occurs. It seems that a realisation of this is implicitly already being used as a criterion within railway safety decision making, where a compromise between utility and state-of-the-art reasoning has been adopted in some cases.

Unless the industry is prepared to adopt a state-of-the-art approach, it will have to expect intense pressure and criticism whenever there is a preventable fatality. This, however, brings us back to the first question: what can be done to remove the sense of culpable negligence, and get the public to understand and approve of the industry's approach to risk (if, perhaps, somewhat modified).

A crucial project is to understand why in detail it is that the public favours the state-of-the-art approach (if indeed it does) and to see how far this is based on false beliefs. The next step would be to look at the issue of public relations from two points of view. One to confront and challenge

any false beliefs that drive the state-of-the-art assumption. For example, it is conceivable that the public may be more prepared to give up the state-of-the-art preference if they felt that the money saved would be used to benefit them in other ways, such as lower fares or greater reliability, rather than going in the pockets of shareholders. The other is to do more to explain to the public what is already being done. I take all this up below.

Part 4. Proposals for Future Work

4.1 Aims of Future Work

The central problem causes a difficulty for the industry because there is a mismatch between the industry's approach to safety and the expectations of certain members of the public. The alternatives appear to be: first, to accept that this mismatch is a fact of life, and continue as before, even though this will create regular difficulties; second, to try to meet public expectations; third to try to bring the public to accept the industry's standards.

It is likely that any realistic strategy may well contain elements of all three. However in some ways the third is the one which promises the most, even if the chances of achieving it are small. As a first step it is an absolute requisite to understand what lies behind the public's existing attitudes to safety. A study is proposed to attempt to uncover this. This will provide information which may allow a public relations exercise to help bring an adjustment in public attitudes. This is the first project described below. To interpret public reaction, a better understanding of culpable negligence may be required. This is the second project described below. In addition, adjustment to public attitudes may also be achieved through changes which do not need the discovery of new information. This is the third project described below. Finally, the analysis in this paper has implications for the nature in which safety decision making should be conducted. This is the final project described below.

4.2 Proposals

4.2.1 Understanding and Reducing Societal Concern

It has been argued that the reason why societal concern is out of joint with individual perception of risk is that there is a belief that in allowing system failures the industry is disgracefully failing in its safety duties. This moral criticism, it is proposed, is based on the assumption that a state-of-the-art approach should be taken to safety. But the deeper reasons why people think this, if they do, are not yet known, although several possible explanations have been conjectured here. These include:

- International Comparisons
- A sense that the record is deteriorating
- Allegations of 'profiteering'
- Quasi-elimination bias

Accordingly consideration should be given to undertaking a study to examine how strongly each of these factors influence public attitudes. It is doubtful that any direct method would work, but various indirect methods may be possible. For example what we might term a sensitivity study may be attempted; where subjects are asked to read certain material before filling in a standard questionnaire. Much more detailed consideration to the design of the experiment is necessary, but for purposes of illustration, groups could, for example, be exposed to:

1 Nothing

2 Tabloid sensationalism, including accounts of accidents, neglect, SPADs, poor maintenance etc.

3 Report on improving safety standards

4 Neutral comparison with road safety and other risks, provided by independent experts

5 Comparison with (eg) Spain, Italy, France and Germany on safety record, technology etc

6 Various combinations of the above.

In addition, consideration should be given to assessing the subjects' responses to different types of risk, to test for quasi-elimination preference, and also to test whether their attitudes are affected by whether the railways are in public or private ownership.

The point would be to test what types of things make individuals more and less sympathetic to current safety policy. Of course there may be other ways of getting this information, and this would be a matter for further discussion.

While interesting in itself, far more importantly this approach may provide guidance as to where the industry should concentrate its future public relations effort to reduce societal concern. For if it turns out that the extreme reaction is based on false assumptions it may be possible to challenge these through well-designed public relations initiatives. The first step is to undertake the psychological study. The second is to commission a public relations expert to review the findings of the psychological study and to consider what prospects there are for a campaign. Both of these steps are recommended.

4.2.2: The Moral and Legal Nature of Culpable Negligence

In the course of this investigation it has emerged that the idea of culpable negligence is critical. However within the remit of this study it has not been possible to review the legal literature and compare it with moral theory, public attitudes, and the current regulatory regime. A short further study may be helpful to attempt to clarify this notion, and cast further light on the issue of when exposing people to known risks can and cannot reasonably be portrayed as culpably negligent. This could form a useful further background to a public relations exercise in which railway risks are compared to other risks.

4.2.3 The public face of railway safety

While the first project requires the discovery of new information before engaging in public relations initiatives, it seems that there are steps that can be taken even now to improve public sympathy for the industry's approach to safety. It is recommended that action is taken under the three heads of press relations; publications and routine communication.

4.2.3.1 Press Relations

Railway Safety documents do not explain how press relations are organised, or who in the industry takes responsibility for it. However it seems clear that relatively little in defence of the industry's record is getting through via the media. This is clearly an area which could be looked at.

4.2.3.2 Publications

Railway Safety's published documents include annual reports [Railway Safety, 2002] and safety plans and reports [Railway Safety 2002b, 2002c], Within this material two main goals are in view: to strive for ever-better safety performance; and to re-assure the public that they should have justified confidence in railway safety.

Inevitably there will be some tension here. Showing that work is taking place to improve safety is likely to send the message that things are not good enough yet. We see this already with Railway Safety's slogan: 'Working For A Safer Railway'. This slogan appears to encourage the thought that the railways are currently unsafe. Perhaps some notion of 'strengthening', 'supporting' or 'maintaining' standards may give out a more positive message.

It may also be more helpful to emphasize that the industry is improving norms of safety and setting ever more rigorous targets rather than coming closer to meeting targets set in stone which it is

currently failing to meet. Improving standards (and meeting those standards) is a more positive message than improving safety. Suggesting that we must never be satisfied is much better than implying that we realise that the railways are not safe.

On the positive side, there are many improvements to display. However, there are so many targets and figures that the overall picture is somewhat bewildering for the public. Thought should be given to providing a simpler form of reporting. This could take the form either of picking a few key targets, or constructing a single 'safety index' which would be some sort of weighted function of the individual targets. On this latter approach, the individual targets would still stand, but this would provide a more digestible summary in which the public could take an interest. It would also allow for annual reporting on yearly targets as well as the final target. This would allow better communication with the public and greater assurance that not only are steps being taken but real achievements are being made. This has, of course, already been done with the Precursor Indicator Model Index, so it would be a matter of further generalising this. Testing may be necessary to see which approach is preferable.

The language of 'zero tolerance', however, seems to me unfortunate, as it is easily confused with two unattainable goals: zero accidents; and introduction of all possible safety improvements without regard to cost. These have the unfortunate effect of reinforcing the assumptions which may underlie the criticism of moral culpability. Ideally this could be reworked.

One of the notable factors in recent public discussions of railway safety is that the public do not accept accident statistics as a guide to risk (cf the discussion of signal value). One interesting feature is that Railway Safety is equally doubtful of accident statistics as a guide to safety, and seeks to study and control 'precursor' factors. This seems an extremely positive step and should be given greater emphasis. It is also an area where there is room for setting new standards as knowledge increases. (On the other hand there is a sceptical note, that certain things that are included as precursors are not really so - certain categories of SPADS - and so where

improvements appear to be taking place they will not actually reduce accidents. If this is so then it ought to be addressed, as it is storing up a future public relations difficulty.)

The idea of setting new standards can be tied in to response to accidents. It may be helpful to do more to highlight the existing commitment to learn from every accident, developing new understandings of risk factors, and seeking ways of reducing them. Also it may also be helpful to emphasise steps being taken to improve accident survivability. Here, while it can humbly be admitted that it is a vain hope to eliminate all accidents, better train design may help to save some passenger lives.

4.2.3.3 Routine Reporting

Finally routine public communication is another area that needs thought. At present passengers are put on notice that there will be planned engineering work, or are told that emergency engineering work is taking place, as reasons to expect delays and disruption. However there is often little communication of the detailed nature of the work or reporting back to the public that it has been achieved or completed. Prominent posters explaining that new track has been laid, or signals replaced, may help convey the industry's efforts to the public. In this respect London Underground seem to do a far better job of routine communication.

In conclusion a number of steps should be considered in detail, possibly by a public relations consultant, any of which may help achieve a better public understanding of the industry's attempts to improve safety.

4.2.4 Safety Decision Making

It appears that, at the present time, there is little public confidence in railway safety decision making policy. The steps suggested above may help adjust public attitudes. However, there

remains the question of whether railway safety decision making needs to be adjusted too. There are three questions to consider:

1. How should railway safety decision making be conducted in the light of current public attitudes, assuming continued use of risk cost-benefit analysis?
2. How can such decision making respond to, or incorporate, changing attitudes, again assuming continued use of risk cost-benefit analysis?
3. Is there an alternative decision making framework which may be more helpful?

Within the remit of this study only a limited investigation of these questions has been possible. Here some recommendations are made for further work in this respect.

4.2.4.1 Joined Up Cost-Benefit Analysis

The current draft railway decision criteria document applies a standard VPF to determine whether a safety measure falls within the ALARP principle. [Railway Safety, 2002d] For an account of how the current figure was derived, see [Maidment, 1998]. It is widely acknowledged that any calculation will have a number of arbitrary elements; for an example of possible criticism see the Appendix herein, which examines the calculation provided by Jones-Lee.)

Certain measures that do not fall within the ALARP principle are allowed or required if certain further conditions hold. These include risks of a special nature; those which combine the possibility of multiple fatalities with high level of industry control of the risk. Here the CBA calculation is based on a VPF with a standard multiplier, said to reflect societal concern.

It is recognised as problematic that actual decisions taken have been much more costly than sanctioned by this method, even using the multiplier. [eg Sharpe, 2001, p. 23] This in itself shows that a consistent methodology has not been adopted, or, at least, some other variable must have been implicitly included.

Using a multiplier for multiple fatality accidents to express societal concern seems to me to be based on a faulty assumption; that societal concern is a form of magnified individual concern. To the degree that it is not, some other approach seems necessary. In a number of documents the extreme commercial implications of accidents are detailed (e.g. [Sharpe, 2001]). Often, though, they are immediately excluded as non-safety issues, but which might, nevertheless be taken note of on commercial grounds (e.g. covering letter of [Marris, 1993]). However this may be an unhelpful division if the commercial costs are largely a result of the public refusing to accept that the industry is meeting its moral obligations to provide for safety. There is good reason, it seems, to use such figures as inputs to a cost-benefit analysis which straddles safety and commercial issues. One may conjecture that, informally, such considerations are already being used. The following may be a reasonable reconstruction of the actual procedure in use:

1. Apply a standard VPF figure.

2. Consider whether the hazard is one especially dreaded. If so add a multiplier to reflect enhanced individual concern. HSE says that currently this applies only in the case of cancer, which is irrelevant to almost all railway safety issues.

3. Consider whether this is a case where risk statistics may mislead about the potential seriousness of the hazard: catastrophic potential; irreversible effects; future generations; inequitable distribution. If so, make appropriate adjustments; in extreme cases this may mean deciding that the hazard is too serious to allow.

4. Consider altruistic concern based on particular concern for any affected individuals; for example that the affected will be children, or another vulnerable group. The use of leaded paint on toys may be an example. In the context of railway safety it is plausible that any multiple fatality accident will engage great sympathy. This may justify a multiplied VPF, although we should note that within the philosophical literature there is disagreement about whether 'other-regarding' preferences should be given any weight at all, for this can lead to discrimination against the unpopular. However, there may be other ways of taking the rights of the vulnerable into account.

5. Consider whether further moral issues can be thought to apply. Can those who impose the risk be reasonably portrayed as reckless or negligent, perhaps even criminally so, even if they meet the ALARP principle? Is it a 'disgrace' that this could happen? If so then there is likely to be societal concern which goes beyond sympathy for the victims. Here it has to be asked whether there are further financial consequences of the hazard occurring: punitive compensation for victims; public enquiries; loss of faith in an industry and consequent falling revenues (and in the case of the railways, passengers choosing to use other and higher risk modes of transport); costs of new regulation; costs of new safety regimes. A factor reflecting this should be brought in as a separate commercial issue. This need not be based on a multiplier of VPF but rather on economic impact analysis of different types of accidents: perhaps first classified according to the type and level of concern to which they give rise. All of this could be factored in as 'the costs of a reputation for poor safety'.

In sum, the current official approach stops at stage 4, and so does not take note of all the costs of an accident. However it is doubtful that the official approach is always followed, and some measures appear to take aspects of stage 5 into account. However this is done in an unsystematic way and the industry still suffers each time a 'preventable' accident takes place. Accordingly a new model for safety decision making should be considered, which makes a proper assessment of the costs in stage 5. It may thus be possible to remain within a cost-benefit analysis framework, but to have a broader understanding of the financial costs of an accident, which include loss of

confidence and reputation, as well as compensation and repair. This requires a full economic impact analysis of different types of accident.

4.2.4.2 Incorporating Public Relations

If one seeks the goal of securing public acceptance of railway safety policy, it may be best to work from both ends: that is, to try to bring the public to change their attitudes as well as modify policy in the light of known attitudes. This means employing a much broader framework in which public relations, in the form of education about safety policy, is seen as part of safety decision making. The strategies discussed above may be part of this exercise. Again it may be possible to remain broadly within a cost-benefit analysis framework, with the costs and benefits of public education being seen as part of a broader equation, and the economic impact of accidents reduced as attitudes improve, although detailed work on this would be necessary.

4.2.4.3 Beyond Cost Benefit Analysis

Finally, for the longer term it may be worth pursuing a project to examine whether any alternatives to RCBA could be viable, in terms of satisfying all parties. This, clearly, would be a project of great complexity, but could start by looking at other regimes for regulating safety, such as product safety regulation, including food and drink regimes, to compare with the HSE approach. It is unknown, however, whether anything of worth could be achieved in this respect, and so this would be a highly speculative project.

4.3.Summary Action Plan

The following research projects are proposed:

a) A psychological study investigating the assumptions behind public criticism of current safety policy. Purpose: to understand current attitudes, in order to know whether or not

they are soundly based, and where they are not, to be in a position to mount a challenge through education and public relations. To be followed up by a public relations exercise.

b) A short study of moral and legal understandings of culpable negligence. Purpose: to enrich understandings of public attitudes.

c) To employ public relations experts to explore how the industry may improve its transmission of information to the public. Purpose: to give the public a greater sense of involvement in the industry and to correct misapprehensions.

d) A project to model safety decision making in a set of broader contexts, including economic impact analysis of different types of accident. Purpose: to investigate the possibility of producing a framework which secures public acceptance

Part 5. Conclusions

On the analysis presented, at the heart of the problem is societal concern floating free of individual concern; people believe that safety is being neglected on the railways, even though they do not think that risks are high. To understand this attitude we need to deepen the analysis presented by R2P2 by paying attention to the manner in which the risk are perceived to be imposed. If it is thought that a risk is being imposed in a way which can be morally criticised - through negligence, recklessness or selfishness - this will lead to a level of societal concern which may not be reflected in individual attitudes to risk. Evidence of neglect is thought to be found in two types of accidents through system failure; first, where it is generally believed that technology to eliminate such failures exists and is in use elsewhere; and second, in accidents which seem to be the result of failures of routine maintenance. This concern is intensified by a number of psychological factors and amplified by the media. People may even say that they are ashamed to live in a country which has allowed its railways to be so neglectful of safety. Bearing in mind that public concern is not with the level of risk, but with the apparent irresponsibility with which these risks are imposed, emphasising the industry's safety record will not, in itself, do much to undo this attitude. Rather steps need to be taken to understand more precisely why the public think that it is scandalous if state-of-the-art technology is not adopted, regardless of cost, and to design public relations initiatives that may counter this sense of scandalous neglect. In the meantime consideration should also be given to joining up safety and commercial considerations in safety decision making.

Appendix: The Jones-Lee studies

Valuation of Benefits of Health and Safety Control: Final Report

Michael Jones-Lee et al

This is a fascinating, though rather extraordinary and revealing study. I shall comment on it at some length for as a study commissioned by the Health and Safety Executive it may have some influence on the further development of the debate, but it seems to me that it should treated with caution, particularly in its calculation of VPF.

The basic thrust appears clear enough; safety improvements cost money. When should we introduce them? Cost-benefit analysis appears to provide an answer. We should introduce safety improvements when, collectively, we are prepared to pay for them. The question then arises of how much we are prepared to pay, and this study looks at how much individuals are prepared to pay to ensure the prevention of a fatality (VPF). It is assumed that the correct methodology is to respond to averaged individual preferences of a representative sample; i.e. what people would be prepared to pay to avoid a risk, or the level of compensation they would need to be paid to accept a risk.

The investigators note that finding that level is fraught with difficulties. Accordingly they have to resort to a very indirect strategy in order to get any sort of robust result. Although one has to have sympathy for the difficulty of their task, one also has to recognise that the exercise relies on so many assumptions at crucial points that it is doubtful that it reveals very much, and the authors seem to claim more than they have established. In fairness it is clear that the authors are aware of the weaknesses of the study; however they do rather blithely brush them aside. But it does not seem to me that the paper does anything like establish a non-arbitrary VPF, based on individual preferences.

It is sometimes said that unless VPF is computed we are left relying on intuition and arbitrary factors in formulating public policy. However bearing in mind that these factors enter into the calculation of VPF, we have a rather spurious contrast.

The railways enter only in the second phase of the study. In the first phase a VPF is established for road safety. In the second phase - a 'relativities study' - a different sample of people are asked to rank risks from roads, the railways, domestic fires and public fires. The result of this, which may confound current expectations, is that the survey suggests that a higher VPF is appropriate for the roads than the railways. This is taken up in the post-Ladbroke Grove follow-up study, which I will discuss later. Here I will comment first under a number of heads:

1. Cost-benefit analysis.

The basic framework adopted is said to be standard cost-benefit analysis. This is taken as a given, and it has to be conceded that at the present time we lack an alternative theoretical framework. I will not here repeat the difficulties with cost benefit analysis explored in Part 3 above. However this section should be read in conjunction with that discussion.

CBA is standardly applied in cases where we cannot rely on the market to provide a satisfactory solution to a question of distribution. Solutions can be unsatisfactory for a number of reasons, but it is well-known that the market will tend to under-supply public goods, relative to demand. In the current case we are looking, in the first instance, at the supply of a public good - road safety - which has elements of private goods (which car you chose to drive, how you choose to drive it) and elements of public goods (e.g. road surfacing, junction design).

However it must be noted that CBA can be used for more than one purpose, or perhaps better, within more than one arena. In particular we must ask two obvious questions:

1. Who will pay the costs?
2. Who will receive the benefits?

In a market in theory at least you pay for what you receive. In the case of public goods it is possible to receive goods you do not pay for, and to pay for goods you do not receive. Consider two cases:

1. A government deciding how much to spend on defence.
2. A government deciding whether to provide a rural train service.

In the first case the benefits accrue to every member of society, and it seems reasonable, at least in the first instance, that all should pay for this, probably through tax revenues. In the second case the benefits accrue only to a small section of society. Should these people be the only ones that pay or should everyone? That is presumably a decision to be made on the basis of public policy, and which although could be informed by cost-benefit analysis cannot take CBA as its only input.

To approach the point a different way, CBA could be used as a form of market research. To provide a rural train service is a large investment, and the agency behind it may want to know if it can be run in an economic fashion. So, on the face of it, it seems reasonable to ask people how much they would pay to use it. In such an exercise the preferences of people who, for example live in cities and would never use the train, are quite irrelevant. What we want to know is whether there is an economically viable ticket price.

On the other hand, we might be interested not so much in market research as in providing a just social policy. Here we might properly wish to explore the preferences of those who live elsewhere, to see whether or not they are prepared to pay a higher tax to subsidise transport for those in rural areas. This may be important information. But there are two caveats to its use:

a) It is not obvious that the preferences of those who do not use the system should be included on the same basis as the preferences of those who do use the system.
b) The preferences of those outside the system should not be treated as decisive. For example if city dwellers are already receiving subsidy for the transport they use, then they may have a moral duty of fairness to contribute even if this goes against their strong preferences.

This study appears to straddle market research and the formulation of social policy. It uses the former as a means of providing a solution to the latter. But given that people will typically pay more for goods and services (and to avoid risks) for themselves than they will for the same goods and risk avoidance for strangers the methodology appears to suffer from a crucial confusion.

To see this, consider how the costs of reducing risks are to be met. Let us concentrate on two options; either we raise the price of travel (the ticket price or fuel tax, say) or we raise taxes to provide a general subsidy. Suppose now we decide only to raise the price of travel. In that case why should the preferences of those who do not travel have any bearing at all on what cost/risk package is acceptable? To the degree that the study is unclear about who the costs will fall on, it is also unclear whether it should have taken into account preferences of people who are not exposed to the risks under discussion. Their preferences are only relevant if we are contemplating public subsidy. (I return to this in comments on Phase 2 and on the follow-up study).

To compound the problem we can note that higher prices and higher taxes are not the only options for paying for risk reduction. A third is to cut other public services, and a fourth is to reduce profits of any private company involved in supply. It seems to me that which option we take is very likely

to influence people's views about acceptable costs, and so this must be made clear at the outset. We cannot expect to get any robust results if we first provide the CBA and then ask who the costs should fall on.

A further, obvious, difficulty is that CBA pre-supposes the controversial assumption that all relevant information can be converted into a single currency - money. This could cause both technical and ethical difficulties. That is, some of the difficulties that are laid out below may be a consequence of translation of everything into financial terms necessary for following cost-benefit analysis.

2. The role of preferences.

The investigators note that the approach bases social decisions on the preferences of individuals affected, and appeal to the idea of democracy to support this approach (3). Yet even within democratic theory this is a controversial view. One line of opposition points to a difficulty with the idea of making preferences basic - they are notoriously unstable and context dependent, as is found in this study. Another argument is that in a democracy we should respect individual ideas of the common good which may not be reducible to sum totals of preferences. This may seem like splitting hairs - can we not understand ideas of the common good simply as preferences for how other people should be treated? Yet there is an important contrast between what people think is right, and what they prefer, which may be something else entirely. That is, people may be prepared to suspend their preferences for the sake of the common good.

Furthermore there is a question about how democratic it is that the preferences of people who are not themselves subject to a risk should be taken into account in evaluating the price to be paid by others to avoid that risk. (This was noted above.)

However even in so far as one takes preferences as basic there is a question whether we should take revealed preferences through actual market choices or expressed preferences discovered through surveys. During the heyday of behaviourism revealed preferences were the standard input, and this was the case for early risk studies (especially in the work of Starr). This approach came under criticism in that real choices are always based on a package of reasons, so we cannot read off preferences for particular elements of a package from actual choices. The alternative expressed preference method does not generally have this difficulty and is capable of delivering much more data. The trouble is that the data is not robust (see below). There is, thus, reason to reconsider a more sophisticated version of revealed preference theory.

3. The failure of direct methods.

The methodology of cost-benefit analysis suggests that one should simply ask people what they would be prepared to pay to avoid or reduce particular risks, and to try to build up from this a coherent account of attitudes to risk. The difficulty is that these questions are very far from people's everyday experience, and so there is no reason to believe that they know how to think seriously about these issues. Accordingly there is no reason to think that we can derive a coherent account. It is rather like asking people how they would spend the money if they won the lottery, and expecting, in immediate response, to get a serious, thought-through answer upon which one could build public policy. The paper itself brings out these difficulties (6-7). However it does not use these difficulties to raise questions about whether this means that people don't really have consistent preferences about risk/cost trade-offs. Rather it seems to be assumed that the preferences are lurking in people's heads somewhere, it is just that the people who have these preferences don't know what they are (yet).

4. The chain of reasoning.

Because of the impossibility of the direct method, an alternative method was devised, and indeed the adopted methodology is an ingenious one. Respondents were asked about their willingness to pay for the certainty of a quick and complete cure for a particular non-fatal road injury, involving only temporary disability, and their willingness to accept compensation for the certainty of sustaining the same injury. This is then extrapolated into an estimated wealth/risk of non-fatal injury rate, based on certain consistency assumptions. The respondents are then asked about gambling the risk of non-fatal injury against the risk of fatal injury. Then finally we can arrive at a posited wealth/risk of death rate.

Unfortunately the details of the responses are not given in the report, but it is clear that the range of valuations provided was wildly variant. It may be that part of the reason for this is that the questions asked are still outside ordinary experience. Few of us have ever been in a position where we have to decide how much we are prepared to pay for the certainty of a complete cure for anything (which was the question they were asked), and again few of us have had to think seriously about acceptable levels of compensation for injuries. We also rarely think explicitly about trading off risks of non-fatal injury against risks of fatal injury. So I suspect that what people say in these circumstances is not robust, in that they may give one answer one day and another answer another. I imagine most people would be unlikely to meet the minimal standards of consistency and regularity assumed in step 2. Indeed, it may be precisely because people don't meet these standards that the indirect method had to be adopted. Whatever figures are derived from this must be treated with great suspicion, unless they tend strongly to cluster together (or differences can be explained on the grounds of e.g. income.) Unfortunately the detailed material is not included in this report, but it seems there is no significant clustering.

Furthermore, it is unclear what the exercise tells us about road safety, rather than attitudes to injury in general. As described it seems a rather accidental feature that the injuries under discussion were said to have been caused by a road accident.

Indeed the justification of the methodology (road safety first; relativities later) was that risks in all but road travel were so small that people could not think clearly about them. However, the actual methodology used does not seem to rely in any way on the greater incidence of risks of road travel. Hence there is no reason why phase 1 could not have been repeated for all four risks, rather than adopting a relativities approach for phase 2. Indeed one could envisage both exercises, as a consistency test for the robustness or generality of answers. This may have been very revealing indeed.

5. Sample size.

Although a larger sample size may have been impractical, it seems strange to give so much weight to a sample of just 164, especially when there are such divergent results.

6. The infinite, outliers and adjustments.

It is noted that as many as 16 (10%) of the respondents gave answers to the standard gamble question which translated into an infinite rate of trade-off between risk of injury and risk of death. These had to be excluded from the calculation of the mean (although they were included in the median). Furthermore 2 respondents gave such high figures that between them they more than doubled the mean. These were excluded in the calculation of the mean in the final version. On this I wish to make two observations:

1. Whether or not it was reasonable to exclude these people, one must recognise that the results of the study, in the end, are dominated much more by the investigators' decision to exclude certain

data than by any other factor (indeed one is tempted to say than by all other factors put together). For example, if instead of excluding these 18 people, they had all been normalised to the value given by the least extreme of them (the 18th person) the mean VPF would have been about £2.8 million. If the 16 excluded had been treated as giving the same value as the 17th the mean VPF would have been about £24 million. With these 18 excluded it comes to about £1 million, which is, of course, much closer to the current standard figure.

2. An alternative to excluding these people would be to say that as so many cannot be accommodated, then there must be something wrong with the methodology. In effect, these 10% deny the basic assumption of CBA - that a financial price can be put on everything, including risks of death.

Finally, it is said that ‘ there is an argument that, if anything, people’s responses to hypothetical WTP and WTA questions may overstate what they would actually be willing to pay’. (13) This may be true but no evidence is cited. Nevertheless, in the light of this ‘argument’ the VPF is then adjusted downwards so that it provides a range incorporating the current DETR figure of £900,000. On the face of it this is entirely arbitrary - outrageously so - and is just one of a number of highly questionable adjustments.

7. Non-normal distribution.

Even with the 18 mentioned above excluded, the mean and the median diverge sharply. We don’t have any further information about the distribution beyond the point that it is ‘right-skewed’ but it is clear it is far from a normal distribution. The answers of the large majority of people lead to a much lower VPF than the mean. This is unexplained, but provides further evidence that people simply don’t know how to think about these ‘consumption decisions’ in a coherent way.

8. From road to rail.

One of the problems discussed in this study is the 'framing problem'; people's preferences are highly context dependent and this includes the form in which the question is phrased or framed. Trying to elicit attitudes to rail safety through the 'relativities' approach falls into this difficulty. We cannot tell from this study how people think about railway risk when they are only thinking about railway risk. Here they have to think about it in the context of risks from domestic fires, public fires and road safety. This may jolt people into the realization that the railways are relatively safe, and it is absurd to treat railway safety in the hysterical way it is treated in the press. Yet if they were to think about it alone, out of this context, they may have very different attitudes.

It is clear that people think very differently about questions when they have to make comparisons which imply hard choices. If you ask people whether more money should be spent on education, almost everyone will agree. But if you ask them whether the health budget, the police force, and so on, should be cut to pay for this, they may then reverse their opinion. This is a well known and quite general phenomenon.

9. Arguments for equality and 'arguments for prioritising'.

It is interesting that the investigators tried to make the survey responsive to factors that are known to influence attitudes to risk. Respondents were presented with an argument for equality: essentially a life is a life wherever it is lost and were also presented with 'arguments for prioritising' - scale, voluntariness and expert knowledge.

However it may be argued that although these are reasons why people do prioritise, they are not always part of a reflective argument in the same way as the argument for equality. Rather, it might be thought, they are to some degree part of a set of possibly irrational background assumptions that people bring to bear in thinking about risks. Bringing them to consciousness may expose them to

the bar of reason, which perhaps they will not pass, or at least not in full force. Very importantly in the current context, although media attention may strongly influence people's attitude to risks, they may be very reluctant to admit that they are so 'shallow' as to be influenced in this way. Similar points can be made about the potential for multiple fatality accidents. Rationally it seems that this should not affect attitudes to risk, but it does seem to have a powerful effect, especially when combined with intensive media attention. Somewhat paradoxically, then, stating these reasons could be a way of mitigating their force, rather than incorporating them positively. Against this, some of these factors seem to have exerted some influence in the study. But it would have been interesting to have seen this controlled for.

Furthermore this aspect of the study suffers from the well-known 'are you a selfish bastard?' problem - people may misrepresent their views so as not to look bad. The investigators point out that although two-thirds thought that their own exposure to a particular risk was a reason for taking it more seriously, about one-third did not. However, although this is what those people said, it is harder to tell why they said it. In this respect the expressed preference approach suffers from the same problem as the revealed preference approach.

10. Human frailty

In the relativities study people were asked to concentrate on the lives saved and nothing else. But as the study recognises (25) it is clear that people could not do that; any policy that saves 10-12 lives is certain to have other beneficial effects, some said. So it is unclear that people were answering, or were capable of answering, the questions put to them.

11. Follow-Up Study

This study was undertaken after the Ladbroke Grove accident, to see how it had affected people's attitudes to railway safety, and in particular if this now became a higher priority in people's minds.

I will comment on two aspects of this study.

12. Quantitative

The point of the study was to see how attitudes had been affected by Ladbroke Grove. To do this the investigators chose to select a different sample of people; essentially those in the London commuter belt. In retrospect this seems unfortunate in one respect, for it introduced a second variable into the study - these are people who make more use of the trains (and this is why they were chosen). If we assume that people are more likely to want to pay more attention to risks that they face, than risks others face, this in itself would be a reason for expecting greater concern.

However, this problem is addressed in the study by isolating the responses of high rail users and low rail users. This unsurprisingly confirms that rail use is a significant factor.
So from this study alone, which did note somewhat increased concern for railway safety, we cannot say whether this is attributable to Ladbroke Grove, or simply to the different constitution of the sample group.

13. Qualitative

The free discussion of railway safety is the first evidence we have of how people view railway safety when not comparing it to other risks. This is particularly interesting as the people concerned are those who have already spent time considering relativities. It is very interesting to see how easily they fall into the attitude that the railways are in a desperate state.

The investigators pick out three themes which dominated the free discussion: the poor state of the system; direct causes of Ladbroke Grove; and moral outrage. Many, although not all, of the general points raised fell under these heads. A complete list of points raised in given in Table 11; which will be very useful in compiling future questionnaires. The general message is that people

report feeling that the railways are less safe after Ladbroke Grove; for example the incident was of a nature that people did not realise could happen.

One theme worth flagging up here is the 'profits before safety' aspect of the 'moral outrage' theme. This will always be a concern for as long as a service aims at making a profit. It may, thus, put a much stricter duty of safety on the operators under these conditions. However the point can be made both ways; there may seem to be less excuse for a public service to be unsafe, for as it is not intended to run at a profit there may be thought to be less constraint on what it ought to spend.

14. Summary

The attempt to derive a VPF for road safety is flawed in a number of ways, and reinterpretation of the data could have led to very different figures.

The relativity study is useful in showing that, in general, there is a low estimation of individual risk from railways. However there was little attention to examining risks from different types of railway hazard, and so the study leaves unexplained why there is such an extreme reaction to some risks.

The most interesting aspect of this exercise is the qualitative discussion within the follow-up study, which appears to contradict the quantitative study and thus re-inscribes our central problem. Although very useful, as it stands the study does not tell us enough about people's attitudes to railway safety to be able to know what measures would do best to help reassure the travelling public. Further research here would be very helpful, and the studies here suggest promising lines. This is elaborated in Part 4 of the main report above.

Bibliography

Adams, J. [2001], *Risk* London: Routledge (first published 1995).

Baier, A. [1986] 'Poisoning the Wells', in [MacLean, 1986].

Flynn, J. Slovic, P. and Kunreauther, H., ed [2001], *Risk, Media and Stigma*, London, Earthscan.

Gibbard, Alan [1986] **'**Risk and Value', in [MacLean, 1986].

Jones-Lee, Michael, *et al* [2000a] *Valuation of Benefits of Health and Safety Control,* HSE.

Jones-Lee, Michael, *et al* [2000b] *Valuation of Benefits of Health and Safety Control: Follow-Up Study*, HSE.

Health and Safety Executive [2001] *Reducing Risks, Protecting People*.

Kasperson**,** R.E., et al [2000] 'The Social Amplification of Risk: A Conceptual Framework', in [Slovic, 2000a].

Leonard, H.B. and Zechkauser R.J, [1986] **'**Cost-Benefit Analysis Applied to Risk: Its Philosophy and Legitimacy', in [MacLean 1986]

MacLean, Douglas (ed) [1986] *Values at Risk*, Totowa NJ: Rowman and Allanheld.

Maidment, D. [1998], *Development of Safety Decision Rules Within the UK Railway*, for Southall Inquiry.

Marris, C. *et al* [1993] *Incorporating Public Perceptions of Risk into Safety Investment Decision Making*, for British Railways Board.

Mitchell, M [2001] 'The Impact of External Parties on Brand-Name Capital: The 1982 Tylenol Poisonings and Subsequent Cases', in [Flynn, Slovic, and Kunreauther, 2001],

Railway Safety [2002a], *Railway Safety Annual Report 2001/02*

Railway Safety [2002b], *Annual Safety Performance Report 2001/02*

Railway Safety [2002c], *Railway Group Safety Plan 2002/03*

Railway Safety [2002d]*Decision Criteria for Railway Group Standards* (Draft).

Railtrack [2001] *Railtrack Railway Safety Case, Version 26*

Sharpe, Andrew [2001], *Railtrack Safety Risk Strategy*.

Sharpe, Andrew [2002] *The Tolerability of Risk: Understanding Ethical Perspectives*, for Railway Safety.

Slovic, Paul (ed) [2000a] *The Perception of Risk*, London: Earthscan.

Slovic, Paul [2000b] 'Perception of Risk from Radiation', in [Slovic 2000a].

Slovic, Paul. [2000c] 'Trust, Emotion, Sex, Politics and Science: Surveying the Risk-assessment Battlefield', in [Slovic 2000a]

Slovic, P., Fisschhoff, B., and Lichenstein, S., [2000a]'Facts and Fears: Understanding Perceived Risk', in [Slovic 2000a].

Slovic, P., Fischhoff, B., and Lichenstein, S., [2000b] **'**Response Mode, Framing and Information-processing Effects in Risk Assessment', in [Slovic, 2000].

Rail Safety and Standards Board Evergreen House 160 Euston Road London NW1 2DX
Reception Telephone +44 (0)20 7904 7777 Facsimile +44 (0)20 7904 7791
www.rssb.co.uk