Really Responsive Risk-Based Regulation

Julia Black and Robert Baldwin

Abstract
Regulators in a number of countries are increasingly developing ‘risk based’ strategies to manage their resources and their reputations as ‘risk based regulators’ have become much lauded by regulatory reformers. This widespread endorsement of risk-based regulation, together with the experience of regulatory failure, prompts us to consider how risk-based regulators can attune the logics of risk analyses to the complex problems and the dynamics of regulation in practice. We argue first, that regulators have to regulate in a way that is responsive to five elements: regulated firms’ behaviour, attitude and culture; regulation’s institutional environments; interactions of regulatory controls; regulatory performance; and change. Secondly, we argue that the challenges of regulation to which regulators have to respond vary across the different regulatory tasks of detection, response, development, enforcement, assessment and modification. Using the really responsive’ framework we highlight some of the strengths and limitations of using risk based regulation to manage risk and uncertainty within the constraints that flow from practical and indeed from the framework of risk based regulation itself. The need for a revised, more nuanced conception of risk-based regulation is stressed.

Introduction

Regulators usually find that they have more do to, and more issues to respond to, than time or resources allow. Partly as a result, many governments and regulators are now developing risk-based regulatory strategies as frameworks for the management of their resources, and their reputations (Rothstein et al 2006; Black 2005). These are collections of strategies that at the very least involve the targeting of enforcement resources on the basis of assessments of the risks that a regulated person or firm poses to the regulator’s objectives.¹ The key components of such assessments are evaluations of the risks of non-compliance and calculations regarding the impact that the non-compliance will have on the regulatory body’s ability to achieve its objectives. In its idealised form, risk-based regulation offers an evidence-based means of targeting the use of resources and of prioritising attention to the highest risks in accordance with a transparent, systematic and defensible framework.

The UK is a jurisdiction that has fully embraced risk-based regulation, at least at the level of exhortation. The Hampton Review of 2005 (Hampton 2005) endorsed the risk based frameworks developed by the Environment Agency, Food Standards Agency, Health and Safety Executive and the UK Financial Services Authority and recommended that all regulatory agencies should adopt a risk-based approach to enforcement. The Government has sought to implement that recommendation across UK regulatory affairs, and all UK regulators are now under a statutory duty to develop and use risk based frameworks for organising all aspects of their regulatory activities, including data collection, inspection, advice and support programmes and enforcement and sanctions (Statutory Code of Practice for Regulators 2007: section 4). The UK is not alone in this move, and regulators have been
developing risk-based frameworks of supervision in a wide range of countries, particularly in the areas of environment, food safety, occupational health and safety, financial services and pensions regulation (Black 2008).

This widespread endorsement of risk-based regulation prompts us to consider how risk-based regulators can attune the logics of risk analyses to the complex problems and the dynamics of real-life regulatory scenarios. Elsewhere we have argued that such attuning requires regulators to be ‘really responsive’ to the ongoing challenges that confront regulators on a daily basis (Baldwin and Black 2008). We, accordingly, look here at the fit between ‘risk-based’ and ‘really responsive’ regulation. By ‘really responsive’ regulation we mean a strategy of applying a variety of regulatory instruments in a manner that is flexible and sensitive to a series of key factors. These include, not merely, the behaviour, attitude and culture of the regulated firm or individual but also the institutional environments in which regulation takes place; the ways in which different control instruments interact; the performance of the control regime itself; and the changes that occur in regulatory priorities, challenges and objectives (Baldwin and Black 2008).

A further prompting to re-examine the implementation challenges of risk-based regulation comes in the wake of the 2007-9 credit crisis and stems from the widespread perception that risk-based regulation, at least in the UK, signally failed to protect consumers and the public from the catastrophic failure of the banking system. This decline in the reputation of the UK’s risk-based approach to financial regulation, together with the experiences of other regulators who have adopted risk based approaches, presses us to examine whether there is a need to apply risk-based regulation in a newly reflective manner and to conceive of it in a more nuanced way. We will argue that a ‘really responsive’ approach offers a framework for re-conceiving risk-based regulation in such a fashion.

This article contends that there are considerable difficulties to be faced in seeking to apply risk-based regulation really responsively but that the payoffs from doing so outweigh any such difficulties. One payoff, it will be seen, is that the ‘really responsive’ framework forces us to move away from the rather too easy vision of risk-based regulation - as a mechanical/quantitative means of solving regulatory problems, as envisaged by the Hampton Report, for example (Hampton Report 2005: Recommendation 1). It thrusts us towards a more complex vision of risk-based regulation - as a particular entry point into, and means of constructing and addressing a core set of regulatory issues. It also compels us to see risk-based regulation not as a free-standing and discrete mode of control but as a strategy that is routinely and often necessarily deployed in harness with a host of other strategies for addressing the challenges of regulatory intervention.

A really responsive analysis also suggests that, once regulators have established their objectives, they should consider how any given regulatory approach comes to grips with the five fundamental tasks that are involved in implementing regulation so as to further those objectives. The tasks are: detecting undesirable or non-compliant behaviour, responding to that behaviour by developing tools and strategies, enforcing those tools and strategies on the ground, assessing their success or failure and modifying approaches accordingly (Baldwin
and Black 2008, 2005). Part of our argument here is that risk-based regulation should be used in a way that takes on board not only the way that risk-based approaches offer a distinctive approach to discharging each of these tasks (with the strengths and weaknesses of that approach) but also the ways that risk-based and other regulatory strategies are likely to interact in different ways across the various tasks.

The key elements of risk-based regulation

The development of risk-based frameworks follows the pattern of many innovations (Black 2005). There have been a few ‘early adopters’ and over recent years the number of regulators adopting some kind of risk-based approach has steadily increased. The later adopters have been directly or indirectly helped by the early exponents. Regulators have communicated the detail of their frameworks and their experiences to other regulators through trans-national networks, such as IMPEL in the environmental context, or by bilateral interchanges (Black 2006). Models of risk-based systems are thus spread across regulators and modified each time. For example the risk-based model of the Australian Prudential Regulation Authority (APRA) was based on that of the Canadian banking regulator, the Office of the Superintendent of Financial Institutions (OSFI) and the UK financial regulator, the Financial Services Authority (FSA). It has in turn been adopted in modified form in a number of different countries, including the Netherlands (for banks and pensions) and Indonesia (for pensions). The Canadian prudential supervision model also formed the basis for the Singaporean financial regulator’s risk-based system. The UK’s risk-based model of financial regulation has been adopted in countries as diverse as France and Columbia. Regulators often ‘mix’ models – so the Portuguese environment regulator, IGAOT, used a mixture of the Irish Environmental Protection Agency’s framework, with that of the Dutch environmental regulator, VROM. The Irish Environment Protection Agency’s framework itself drew on that of the Environment Agency for England and Wales and its food hygiene framework draws on that of the Food Standards Agency’s Code of Practice for England (Black 2008).

The frameworks vary considerably in their complexity. However all have a common starting point, which is a focus on risks not rules. Risk-based frameworks require regulators to begin by identifying the risks that they are seeking to manage, not the rules they have to enforce. Regulators are usually over-burdened by rules. They cannot enforce every one of these rules in every firm at every point in time. Selections have to be made. These selections have always been made, but risk-based frameworks both render the fact of selection explicit and provide a framework of analysis in which they can be made.

The frameworks themselves have five common core elements. First, they require a determination by the organisation of its objectives – of the risks ‘to what’ that it is concerned
with. Secondly, they require a determination of the regulator’s own risk appetite – what type of risks is it prepared to tolerate and at what level. This can be an extremely challenging task for a regulator. In practice, a regulator's risk tolerance is often ultimately driven by political considerations. All regulators face political risk, the risk that what they consider to be an acceptable level of risk will be higher than that tolerated by politicians, the media and the public, and that the uncertainties that they face will be unrecognised and / or intolerated. Political risk is in practice a critical element in any risk-based system, as discussed below.

Second, risk-based frameworks involve an assessment of the hazard or adverse event and the likelihood of it occurring. Terminology varies: food and environmental regulators tend to talk in terms of hazards and risks; financial regulators talk in terms of impact and probability. But in general, two broad categories of risk are identified: the inherent risks arising from the nature of the business’s activities and in environmental regulation, its location; and management and control risks, including compliance record. The methods by which management and control risks are combined with or offset against inherent risk scores varies but, broadly speaking, the regulators are concerned with the effect of management and controls in either exacerbating the inherent risk or mitigating it. Although the terminology of risk is used throughout, in practice, regulators will operate in quite differing conditions of uncertainty. In some scenarios there will be high numbers of incidents from which data on their probabilities of occurrence in different situations can be assessed but, in other circumstances, the regulators will be dealing with low frequency events, from which reliable probabilistic calculations cannot easily be drawn or with conditions of uncertainty, where the risk is inherently insusceptible to probabilistic assessment.

Risk assessments may be highly quantitative (as in environmental regulation) or mainly qualitative (as in food safety regulation in the UK, or financial supervision more generally). Quantitative assessments involve less individual judgement and in environmental regulation are often performed by the firm themselves (as in England and Ireland) or can be contracted out by the regulator to a third party (as in Portugal). Qualitative assessments allow for more flexibility and judgement, but critically rely on the skill and experience of regulatory officials who are making the subjective judgements. Third, regulators assign scores and/or ranks to firms or activities on the basis of these assessments. These scores may be broadly framed into three categories or traffic lights (‘high’, ‘medium’ or ‘low’) or there may be a more granular scoring system, with five or more categories (the UK Financial Services Authority has fifteen different categories, for example). Where numerical scores are used, these will often operate as shorthand for more complex underlying judgements and they may conceal hesitancies and qualifications in the confident exposition of the number itself. (The expression of subjective judgements in numerate form can also lead observers to misconstrue all risk based systems as purely quantitative, whereas in practice their character can vary quite considerably between regulators.) For the most part, assessors do not indicate whether they think the risk score is likely to increase or decrease over time - though one notable exception is the Canadian banking regulator, OSFI, which requires its supervisors to indicate the ‘direction of travel’ of the risk, and the time period over which they think this will occur.
Fourth, risk-based frameworks provide a means of linking the organisation and of supervisory, inspection and often enforcement resources to the risk scores assigned to individual firms or system-wide issues. In practice, resources do not always follow the risks in the way that the framework would suggest, but resource allocation remains a key rationale for their development.

Despite these common elements, no two risk-based systems are identical in their form and they often differ significantly in their operation - even if they happen to have certain similarities in form. Some of these formal and operational differences stem from the regulators’ often widely differing remits and their locations within the institutional structures of their governments. But differences also reflect strategic choices and, as such, can be revealing. Risk-based frameworks are not neutral, technical instruments. Each aspect of a risk-based framework involves a complex set of choices. They require decisions by the regulator regarding such matters as: the risks it will identify as requiring attention; the indicators and methods it will use to assess those risks; where it will prioritise its attention, and where it will not, and, ultimately, of political risk: what level of risk or failure the regulator is prepared to accept - or at least thinks it can withstand.

The Elements of a Really Responsive Approach to Risk-Based Regulation

The ‘really responsive’ approach, to recap, carries two main messages. The first of these is that, in designing, applying and developing regulatory systems, regulators should be attentive and responsive to five key factors: the behaviour, attitudes and cultures of regulatory actors; the institutional setting of the regulatory regime; the different logics of regulatory tools and strategies (and how these interact); the regime’s own performance over time; and finally, changes in each of these elements.

These five factors are chosen because they encapsulate the central challenges that regulators face and which must be risen to if they are to achieve their objectives over time. The behaviour, attitudes and cultures of regulatory actors are considerations that require responsiveness because the motivational postures, conceptions of interests and cognitive frameworks of regulated firms (and regulators) heavily influence the regulatory relationship and the regulator’s capacity to influence regulate behaviour (Oliver 1991). The institutional setting of the regulatory regime has to be taken into account because the position that each organization (regulator or regulatee) occupies with regard to other institutions can have a critical effect on the actual and potential operation of regulation. The actions of a regulatory agency, for instance, are strongly shaped by the distribution of resources, powers and responsibilities between that body and other organizations, including those who oversee it (Scott 1995; Powell and DiMaggio 1991).

The interactions of different regulatory tools and strategies have also to be responded to because they impact pivotally on regulatory performance. Most regulators use a wide variety of control tools and strategies but these often have divergent logics – they embody different regulator to regulatee relationships and assume different ways of interacting. Thus,
command and sanction-based instruments operate on very different understandings to educative or economic incentive systems of control. There may be harmony or dissonance between these tools and strategies – so that, for instance, applying sanctions on a deterrent basis may undermine a concurrent strategy of ‘educate and persuade’ by killing regulator to regulate communications. It is, accordingly, essential for the really responsive regulator to manage tool and strategy interactions (Gunningham and Grabosky 1998; Black 1997).

Being sensitive and responsive to the regime’s performance is also of crucial importance to any regulator. If regulators cannot assess the performance of their regimes, they cannot know whether their efforts (and budgets) are having any positive effect in furthering their objectives. Nor can they justify their operations to the outside world. If they cannot modify and adapt their operations and strategies in the light of performance assessments, they will be saddled with poor delivery and incapable of dealing with the new challenges that all regulators are confronted by.

Finally, the ‘really responsive’ approach holds that sensitivity to change lies at the heart of acceptable regulatory performance. In virtually all sectors, regulatory challenges are in a state of constant shift. Thus, for instance, new risks and risk creators emerge or are recognized, and uncertainties can harden into risks, for example as events occur, knowledge develops, technologies and markets change, institutional structures are reformed, political and legal obligations alter, and public expectations and preferences mutate. If regulators cannot adapt to change, they will apply yesterday’s controls to today’s problems and, again, under-performance will be inevitable.

As for the exhaustiveness of the five key factors of the ‘really responsive’ approach, it can be argued that regulators who attend to the above matters will have cause to come to grips with all of the main challenges that are identified by the prevailing theories of regulatory development. Regulators are thus called on to take on board: the importance of divergent interests (be these public, private/economic or group); the significance of variations in cultures, values, ideas, communications regimes and control systems; and the impact of intra- and inter-institutional forces (Baldwin and Cave 1999; Morgan and Yeung 2007). They will also be sensitive to the ways in which regulatory challenges and interactions vary across regulatory issues and tasks – as is the message of ‘regulatory space’ theory (Hancher and Moran 1989).

The second message of ‘really responsive regulation’ is, as noted, that regulatory designs, developments and operations should take on board the way that regulatory challenges vary across the core tasks that regulators have to carry out, both with respect to individual firms and in developing strategies more generally – namely: detecting undesirable or non-compliant behaviour, responding to that behaviour by developing tools and strategies, enforcing those tools and strategies on the ground, assessing their success or failure and modifying them accordingly. These tasks of implementation are central to any regulatory strategy. In emphasizing them, we do not suggest that the logically prior task of determining objectives is straightforward or irrelevant, nor, indeed, that in practice the task of identifying objectives is wholly separable from issues of implementation: regulators like other decision makers can in practice define their objectives with reference to what they think they can achieve rather than (or as well as) what they would hope to achieve in an
ideal world of infinite resources and no institutional constraints. Further, we recognize that few organizations have complete freedom of action and that, in introducing new strategies, the organisation’s embeddedness has to be recognised and taken on board, both by the organization itself and by prescribers of those strategies, be they academics or politicians. Here, the case for focusing on and looking ‘across tasks’ is that there is good evidence that the work to be done to achieve real responsiveness will vary significantly from task to task and that it would be a mistake to think that a strategy that works in relation to, say, the detection of non-compliers will prove as effective in relation to the securing of compliance or the assessing of performance (Baldwin and Black 2008).

The special challenges of applying a risk-based regime in a ‘really responsive’ manner should now be considered.

i) Responsiveness to the behaviour, attitude and culture of firms

Regulators generally need to take on board the organisational capacity, past behaviour and attitude of the regulated firm if they are to achieve their mandated objectives. For risk-based regulators, such factors will routinely be considered by making reference to risk scoring formulae when targeting regulatory interventions at the highest risk creators. These formulae are often only loosely based around the common risk model that risk equals probability times impact. Impact analyses advert to the quantum of the adverse impact that the firm’s behaviour might have on the achievement of regulatory objectives – they will look, for instance, to the size of the potential environmental, food safety or financial harms that might be caused by the particular firm’s type of operations. They can also refer to the nature of the harm to be suffered: for example the Office of Fair Trading ranks the impact of home debt collection as high because of the nature of the harm that can be caused by abusive practices, rather than because of their systemic impact. Probability calculations will look to the chances that the firm will cause such a harm to occur. Such calculations, are often based, inter alia, on the past and current behaviour of regulated firms or individuals. Thus, poor compliance histories and underperforming risk control systems are often reflected in high risk scores.

Risk-scoring accordingly appears highly compatible with the really responsive approach and its demanded advertence to the capacities, cultures and understandings that operate within regulated organisations. The quality and character of management and their risk controls is often a key mitigator or exacerbator of inherent risks and there is, on the face of it, no reason why probability calculations cannot take on board motivational postures as reflected in such factors as commitment to or accommodation of the regulatory agenda (i.e. the firm’s degree of acceptance that it has to change its behaviour to suit some wider public policy goal); amenability to supervision/capitulation to the regulatory authority; resistance, game playing and disengagement (e.g. FSA 2006; Oliver 1991).

Certain challenges do, however, arise when it is sought to incorporate behavioural and cultural matters into risk-based assessments. A first of these flows from a key difference between risk-based assessments and compliance-based assessments. Risk-based assessments are, at their heart, forward-looking. They attempt to assess the risks of the firm on a dynamic, on-going and future basis rather than seek to capture the state of the firm at the particular
point in time when the inspection or supervision visit occurs. Different types of risk-based systems are better equipped to do this than others. The system used by environmental regulators, for example, focuses only on past compliance records, assigning scores on the basis of past infractions, or whether a particular management system is in place. More subjectively-based systems are intended to incorporate a more responsive and dynamic assessment. Thus, in October 2009 Hector Sants, Chief Executive of the FSA looked back at the credit crisis and stated: 'The key lesson learnt is that the FSA must be proactive and not reactive to the management of risk. This is primarily a matter of judgement.' (Masters 2009)

Such prospective subjectivity can be viewed as a strength of risk-based regulation but it can present difficulties. Discretionary decision-making has to be incorporated within a risk regime in a manner that ensures consistency across evaluators and this can present managerial problems for regulators. For many regulators it is hard to strike the right balance between ensuring that staff apply a healthy level of sophisticated and informed qualitative judgement in their risk assessments and controlling those judgements for the sake of consistency. Further dangers are that the processes of overseeing discretionary decision-making can prove excessively costly in staff time and resources and that centrally-administered controls, checks and structuring procedures can render the agency slow to respond to changes in the regulatory challenges that they face.

The Hampton Report recommended that risk based approaches should not be used just for assessing the risks that firms pose to the regulators’ objectives but across all regulatory tasks (Hampton Report 2005). Two central difficulties have to be faced, however, in using risk scoring as an organisationally-sensitive guide to the execution of all regulatory tasks. First, the ‘really responsive’ approach reminds us that regulators and firms can interact quite differently across the various tasks of regulation and that assessments of cultures and attitudes may vary within firms and across tasks. Some parts of a firm may be more amenable to regulatory intervention than others. Further, even where the firm’s response is not internally fragmented, a firm may prove to be highly resistant and uncooperative in relation to the regulator’s detection work but it may be very compliant once its behaviour is placed at issue (e.g. it is secretive and defiant on disclosure but ‘comes quietly’ when its errant ways are discovered). The challenge here is to develop risk analyses that are sufficiently fine-grained to accommodate such variations rather than to settle for using a crude across the board mode of evaluation.

The second complication that arises is a more general difficulty in using risk analyses to guide all regulatory operations. It is that such analyses may prove far more helpful in relation to some tasks than others and it cannot be assumed that they will provide consistent messages with regard to different tasks. Thus, risk scoring may provide a very ready basis for detecting high risk actors, but it may offer far less assistance in identifying the modes of intervention that are best attuned to securing compliance. A firm may, for instance, be given a high risk score because of the inherent risks it poses and because its management’s slack attitudes are reflected in a failure to manage risks well. This risk score may indicate the degree of urgency with which some regulatory intervention has to be made but the firm’s high risk score does not indicate, in itself, whether the best way to reduce the risks posed by the firm is to use a command and control regime applied in, say, a deterrence fashion or whether
an incentive-based, educative, escalating sanctions or disclosure strategy would prove more effective. The kind of intervention required, may at best be loosely linked to the level of risk that the firm presents.

The major determinant of the optimal style of intervention is liable to be revealed by an analysis of the likely responsiveness of the firm to different stimuli – and this may involve a departure from an overly-rigid risk-based system and a drawing on other theories, such as ‘compliance’, ‘deterrence’, responsive regulation, problem-centred and other approaches to fit the context (see generally Baldwin and Cave 1999: Chapter 8). Two firms with similarly high risk scores may, for instance, be respectively well-intentioned and ill-informed or ill-intentioned and ill-informed. The former may respond well to an educative programme and the latter is unlikely to. The former does not need to be met with a punitive threat, the latter may have to be. How best to deal with these two firms is not readily identified by reference to a risk assessment system.

In practice, regulators vary in the extent to which their risk-based assessment systems are linked to a particular policy of intervention. Some, such as Australia’s APRA or Canada’s OSFI, have a close link. An assessment that a firm poses a high risk to the regulators’ objectives will mean that a particular intervention policy is adopted. More frequently, however, regulators have no particular link between the assessment and the nature of intervention or enforcement policies to be adopted.

Another difficulty arises if the same risk-assessments are used as guides to the execution of different regulatory tasks. Risk assessments may offer very ready bases for identifying high risk actors but they will only serve to establish the foremost priorities for the application of enforcement resources if risk-based regulation is used simply to attack the highest risks – rather than to maximise the furtherance of statutory objectives. This point is made by contrasting two hypothetical companies: Millco and Scatterco. Millco presents, let us say, 80 units of risk but it is cheap to regulate because its managers are well organised and amenable to instruction and it occupies one site. The risks it presents can be reduced to 20 units by the application of £5000 of enforcement resource. Scatterco presents 90 units of risk but it is expensive to regulate because its managers are disorganised and not amenable to instruction and it occupies a large number of sites. The risks Scatterco presents can be reduced to 70 units by the application of £5000 of enforcement resource. The given resource of £5000 buys a risk reduction of 60 units in the case of Millco but 20 units for Scatterco.

How should regulatory resources be applied to Millco and Scatterco? The example shows that simply targeting intervention at the highest risk presenter is a very inefficient way of reducing potential harms and that it is not possible to use risk scoring as a driver of action without misapplication of resource. What is arguably needed for a really responsive risk-based regime is a linking of the risk scoring system to an ‘amenability analysis’ that looks at the attitude, nature and organisation of the firm as this relates to: a) the mode of intervention likely to prove most efficient in furthering statutory objectives and b) the anticipated benefits yielded per unit of regulatory expenditure.
A central issue here is whether a risk-based regulator should be responsive to attitudinal and cultural matters by taking levels of amenability on board when risk scoring for the purposes of identifying targets for intervention. There are alternative approaches. If it is thought advisable to separate the identification of highest risks from issues of enforceability, then amenability - or its absence - could be treated as a risk to the agency’s achieving its statutory objectives - but as a different kind of risk to the primary risk that is subject to risk scoring for the purposes of targeting. Thus, ‘primary risk’ would be the risk that, say, Scatterco presents to the agency’s objective to protect air quality – it would be scored in the normal way by looking at the quantum of inherent risk and the likelihood of a harm’s occurring – as mitigated by the firm’s risk management regime. The secondary risk would be the risk that the regulated concern’s behaviour cannot be modified in a manner that renders the primary risk acceptable. Another approach to these challenges would be to treat amenability, again, as separate from ‘primary risk’ but to treat amenability deficiencies not as a secondary risks but as challenges (to modify behaviour so as to render primary risk acceptable) that are seen in terms other than risk – as, for example, issues of political or policy judgement.

For the purposes of this article we are agnostic on whether amenability issues should be seen as secondary risks or as policy challenges. On, however, the case for dealing with primary risk in a way that separates it from secondary risk (or ‘amenability challenges’) this can be evaluated by considering the argument for incorporating the assessment of secondary/amenability risk into the primary risk scoring system.

What can be restated here is that most existing risk based scoring systems do take on board the quality of the firm’s risk management system – at least where corporate arrangements are sufficiently stable to allow evaluation. The weight that is attached to such managerial assessments within risk scoring does, however, vary across regulators and the designs of their particular risk-based systems. The England and Wales Environment Agency’s framework gives such assessments relatively little weight, but the Portuguese environmental regulator’s scheme applies a multiplier of three to the ‘management and control’ score as it wants to incentivise firms to improve their risk-management systems and thus lower their risk scores. Management and control assessments also have a significant impact in the risk scoring systems of many financial regulators since these can lower or raise the ‘net risk’ of the firm. Whatever weighting is given to ‘management and control’ within a given regime, however, the general practice is to lower the risk score overall where the regulator has confidence in the management team’s ability to control relevant risks.

Assessments of firms’ abilities to manage their risks are not, however, the same as evaluations of their amenability – their likely responsiveness to regulators’ interventions. Some firms, for instance, may be competent and serious risk managers but highly resistant to ‘interference’ from regulators. Scoring systems can, nevertheless, be developed to incorporate such amenability assessments. The Portuguese environmental regulator, IGOAT, for example includes in its assessment of management and control an assessment of the firms’ interaction with the regulators and uses this as an indicator for the firm’s amenability to intervention
(Black 2008). In the absence of such a focused assessment, however, it is dangerous to treat an evaluation of the firms' ability to manage its risks, or its attitude to risk management, as an indicator of amenability.

Where amenability assessments are undertaken, there is a strong argument for separating the assessment of amenability from risk assessment. If amenability is included in the overall risk score, this will conflate the risk that the firm presents with the ease with which its behaviour can be influenced. If amenability is assessed separately, regulators can choose whether to prioritise high or low amenability firms for intervention. That choice would be masked if the amenability assessment was bundled in with the primary risk score and an important policy dilemma would be side-stepped. Thus, if firms with low amenability were given a high risk score because of their lack of amenability, this would establish an inverse relationship between priorities for intervention and those areas where risk reductions could be achieved at lowest cost. It would increase the intervention priority given to those firms that would be the most difficult and expensive to bring into compliance. If, alternatively, firms with low amenability were given a low risk score because of their lack of amenability (de-prioritising firms who were unresponsive to influence) this could incentivise resistance to regulation and would reduce the intervention priority given to firms who might otherwise score as high risk. Separating the amenability analysis from the overall risk score thus avoids a problematic conflation and brings two advantages. First, it allows opportunities for lower cost risk reductions to be identified and, second, it enables regulators to make clear decisions on the balance that they want to set between targeting resources at the highest risks and focussing attention on lowest cost risk reductions and efficiency of resource use. This latter clarity of decision-making will be important in many contexts because the optimal balance between ‘highest risk’ and ‘highest risk reduction’ targeting will vary from risk to risk and may be affected by such matters as the regulatory objectives and the political constraints within which a given set of intervention activities takes place. Striking that balance may, moreover, involve political contention.

Similar arguments can be made for keeping the risk scoring of ‘primary risk’ separate from the risks and challenges that flow from the matters to be discussed in the following sections – namely from the institutional environments of regimes; the interactions of regulatory controls; the needs to assess regulatory performance; and the difficulties of coping with change. The discussion of this point will not accordingly be rehearsed in those sections.

Before concluding on the challenges of an amenability-sensitive approach to regulation, however, it should be noted that information on amenability will not always be readily available. It is not, for instance, routinely possible for a regulator to calculate the amount of risk reduction that an intervention strategy will in fact produce in a regulatee. In the Millco and Scatterco example it was assumed that the regulator can calculate the potential risk reduction that will be achieved from a given regulatory effort. In practice, however, this will not always be the case. Such a calculation might be made if the regulatory task involves a well-known regulatee and a high number of similar instances of risk reduction (as might be the case in relation to, say, certain occupational health and safety risks) but it is unlikely to be possible where the risk is one that is low frequency and idiosyncratic and where contacts with the regulatee are infrequent. In such circumstances, where a regulator is dealing with low
frequency events, or indeed with uncertainty rather than risk, it will be extremely hard to predict the risk reducing return from the regulatory input.

To summarise, what the ‘really responsive’ framework suggests is, first, that risk scoring, in relation to any task, should take attitudinal matters on board, second, that attitudes (and their impacts on the risk framework) may vary across regulatory tasks and, third, that astute regulators will be clear about the degree to which any particular regulatory task can and should be guided by a risk scoring system. Further messages are that risk assessments relating to different regulatory tasks may be best kept separate and that risk-based regulation cannot stand alone. Its methods will often require supplementing by additional strategies and judgements since risk analyses offer little assistance in relation to many of the familiar problems that regulators face – such as how best to secure compliance, or whether it is best to regulate in a precautionary or intervention-averse manner. Even on a relatively simple matter such as the identification of the largest risks, risk-based regulation per se tells us only a limited amount about the costs of securing risk reductions or whether objectives are best furthered by targeting analyses at individual risk-creators or particular types of risk or particular industrial sectors or certain systemic risks or some other focus. All of this is not to say that risk-based regulation is inevitably misguided, it is to argue that, although it is often conceived of as a straightforward approach, it is a complex and nuanced strategy that is often dealing in situations of uncertainty rather than risk, although uncertainty is often unhelpfully obscured by the dominant language of risk, and which has to be used with the awareness of its limitations that a ‘really responsive’ viewpoint offers.

**ii) Responsiveness to institutional environments**

A really responsive approach to risk-based regulation emphasises the degree to which regulatory interactions, processes or outcomes are shaped by the institutional environments of both regulatee and regulator. These environments are constituted by the organisational/regulatory, normative, cognitive and resource-distribution structures in which these actors are situated (Scott 1995; Powell and DiMaggio 1991). The actions and decisions of organisations and individuals (both regulators and regulatees) are thus structured by the norms regulating their conduct, by the senses of appropriateness of actions, of understandings of how the environment operates and by the distribution of resources between themselves and others with whom they interact. For regulatees, this includes the market and in particular its value chain. For both regulatee and regulator, it includes other regulators within a decentred or polycentric regulatory regime. A really responsive approach, accordingly, demands the recognition that when regulators apply risk-based approaches, they take into account the challenges and limitations imposed by institutional environments which are both internal and external to regulatory organisations and regulatees and which involve these actors in active as well as passive roles. Two key lessons of the financial crisis, for example, are that risk assessments need to move from their micro-level, firm specific analyses of risk in two ways. They need to incorporate assessments of the impact of macro-economic factors on financial institutions (the ‘outside-in’ assessment) and, further, that they need to couple their micro-level, firm specific analyses of risk with an analysis of risks arising across the financial system as a whole. Part of this requires regulators to incorporate assessments of
the impact of the financial institution itself on the rest of the financial system (‘inside-out’ impacts). Evidence to date suggests that financial regulators are getting to grips with the former more quickly than they are the latter.

Focussing on regulatory organisations, the norms and cultures of these bodies are critical to their operations. This presents special challenges for risk-based approaches that seek to incorporate qualitative, and thus subjective, judgements by regulatory officials into their risk scoring regimes. The formal design of the risk-based assessment system can structure that judgement but it cannot ensure that valid subjective judgements are made in practice. As some of its longest standing practitioners emphasise, risk assessments are an art, not a science. Qualitative risk assessment systems, as noted above, place significant reliance on the quality of regulators’ judgements - which may not be sufficient to the task – and the control of such judgements raises resource and responsiveness issues. In addition, regulators, like others making risk assessments, are prey to cognitive biases and to over-confidence, which will affect their perceptions and judgement. This can have a significant impact on the assessment. Some regulators have found, for example, that their supervisors are particularly poor at assessing the quality of management and controls. Back-testing of assessments reveals that they routinely over-estimate the quality of these controls, by a factor of about 30%, suggesting greater attention needs to be paid to how supervisors’ exercise their judgements in assessing risks (Black, 2008).

The broader institutional, including political context, is also critical to the performance of a risk-based regime yet such regimes may experience special difficulties in dealing with these matters. To return to the example of the failures of the UK’s financial services regulatory regime in the period up to the credit crisis, it could be argued that at least some of these can be put down to key aspects of the institutional environment within which the regulators worked – notably the way that the UK Government’s ‘light touch’ regulatory philosophy shaped regulatory interactions and understandings about the appropriateness of regulatory demands; the degree to which domestic regulators placed faith in controls by other national regulators to control globally interconnected markets (a reliance which was itself in part mandated by the EEA passporting system for financial institutions which placed responsibility for supervision of overseas branches on home state regulators and whose limitations were evidenced by the impact of the collapse of the Icelandic banks on UK deposit holders); and the extent to which domestic regulators considered themselves constrained by regulatory competition within the international institutional environment (HM Treasury 2009; Turner 2009; Tett 2009).

A further difficulty for risk-based regulators can arise when their powers are fragmented or shared. Thus, another factor that may have reduced the effectiveness of the UK’s financial services risk-based regulatory regime in the lead up to the credit crisis was the way in which regulatory powers were distributed between the Treasury, Bank of England and the Financial Services Authority. This arrangement exemplifies the common position in which many risks and social or economic problems are controlled by networks of regulators rather than bodies enjoying the luxury of a regulatory monopoly – networks in which regulation is ‘decentred’ rather than simple and focussed (Black 2001). If attention is paid to institutional environments, the challenges of working within networks have to be taken on board. For
risk-based regulators these challenges may prove considerable and it may be necessary to cope with: divergence between the various networked regulators’ aims, objectives and institutional environments; variations in regulatory cultures; differences in capacities, skills and resources; and varying capacities to modify their operations (Bardach 1998; Sparrow 2000; Kickert, Klijn and Koppenjan 1997; Sullivan and Skelcher 2002; Chisholm 1989).

If, furthermore, we look across the different tasks of regulation, we can see that institutional environments arguably impact on the discharging of all of these – and not necessarily in the same ways. This is a further challenge for risk-based regulators. To stay with the example of UK financial regulation prior to the credit crisis, the detection work of the UK financial regulators was arguably impeded by institutional understandings – for instance that the credit ratings agencies, the banks and the markets could be relied upon to ensure that the sale of collateralised debt obligations and other securitisations produced benign risk-spreading rather than a deep-seated systemic threat to financial confidence. It has also been argued that the institutional environment impeded detection work because it created an expertise imbalance whereby the ‘rocket scientists’ in the investment banks were able to devise bundled products and models to assess their own financial risks (and those of the products they were dealing in) that the regulators were ill-equipped to evaluate in depth (Tett 2009; US General Accounting Office 2009: 6; Haldane 2009; Gerding 2009).

Response development by UK financial service regulators – the devising of new powers and control tools – was constrained by governmental concerns that domestic regulators did not reduce the UK’s regulatory competitiveness and lower the UK’s position in tables of good places to do business. These concerns have legal expression. The Financial Services Authority’s legal mandate requires it to take into consideration ‘the international character of financial services and markets and the desirability of maintaining the competitive position of the United Kingdom’ (Financial Services and Markets Act 2000, s.2(3)(e)). The UK government repeatedly lauded the UK’s ‘light touch’ approach to regulation, which was portrayed as a key weapon in the battle for business between New York and London. It is notable that in their evidence to the Treasury Select Committee that both the chairman of the FSA and the Governor of the Bank of England emphasised the political and market hostility that they would have faced in ‘stopping the party’ and requiring banks to rein in their activities (Lord Turner’s evidence in response to Q 2145 HC Treasury Select Committee 25 Feb 2009; Mervyn King’s evidence in response to Q 2354 HC Treasury Select Committee 26 Feb 2009).

Another institutional constraint on response development within risk-based systems can occur when risks spread across jurisdictions or levels of government so that new rules require levels of institutional co-operation that are difficult to bring about. Thus, it is widely accepted by the G20 governments, including the UK Treasury, that a key contributing cause of the credit crisis of 2007-9 was the failure of national regulators to respond in coordination with other national regulators and supra-national regulators not only to the excessive risks being taken by some individual firms, but to the problems of global system – wide risk and the aggregate effect of individual risk positions being taken by firms whose cumulative effect was to exponentially increase fragility of system by developing global monitoring systems and regulatory structures (H M Treasury, 2009: para. 3.1; G20 2009).
Again with reference to UK financial regulation prior to the credit crisis, enforcement activities were arguably muted by a political and governmental context that was unsympathetic to interventions other than those consistent with light touch regulation. Assessments of regulatory performance were hindered by the regulators’ confidence that few systemic risks flowed from securitisation because the market institutions were managing these and modifications of regulation were held back not only because of incomplete performance assessments but because the regulators knew the Government to have no appetite for more rigorous controls (Lord Turner’s evidence in response to Q 2145 HC Treasury Select Committee 25 Feb 2009).

Some commentators may take issue with some of the above points about regulatory deficiencies and the credit crisis but, for the purposes of the current discussion, the above account serves to demonstrate the value of a really responsive approach as a means of revealing the special challenges that a risk-based regime faces. Noteworthy among these are, first, the need to recognise that regulators are often not dealing with risk, but uncertainty (see also Gray 2009) Second, whether they are dealing with risk or uncertainty, there is a frequent need to depart from the promise of a technical solution to regulatory issues and to incorporate qualitative judgements within analyses. Third, risk based regulatory frameworks are inevitably subject to – and, perhaps, especially vulnerable to - moulding by their institutional and political environments.

On this last point, particularly, the credit crisis revealed a critical aspect of the risk-based regulator’s institutional environment: its political licence to operate. Others have noted that firms need a ‘social licence’ to operate (Gunningham, Kagan and Thornton (2004). We suggest that regulators need a political licence. They need political support if they are to act aggressively against firms, and not just in financial markets (although in financial markets they need considerable support to act against ‘bubbles’ as well). Moreover, risk-based systems themselves carry significant political risk for regulators. Risk-based regulation requires regulators to prioritise, and thus to decide to commit fewer resources to fulfilling its legal mandate with respect to some firms, or risks, than others. The regulatory reality of selection clashes with the political, and civil, expectation of universal protection and the transparency of the selection process can render the regulator politically vulnerable. Transparency has a downside as well as an upside and that downside comes especially to the fore in two kinds of circumstance. First, particularly high political and reputational risk are run when things go wrong in firms or areas of the market which the regulator has decided are low risk and thus to which it has given little attention. Second, in times of crisis risk-based regulators can find that they lose their autonomy to set their own levels of risk-appetite.

**iii) Responsiveness to the logics of control instruments**

As noted above, risk-based regulation usually operates in combination with other strategies. Really responsive risk-based regulation accommodates this reality by paying attention to the ways in which the logics of different regulatory strategies and tools can interact (Gunningham and Grabosky 1998; Black 1997). Different regulatory strategies can have different logics (e.g. of punishment or restoration or rehabilitation or ‘professional’ or ‘commercial’ logics). Particular logics involve distinctive relationships and modes of
conversing with regulated parties – a punitive message, for instance, will be framed and received differently from a rehabilitative message. They involve different understandings regarding the nature of behaviour or of an institutional environment, and in turn have different preconditions for effectiveness. These are based on different assumptions, value systems, cultures and founding ideas so that mixing logics involves distortions and failures of contact.

Really responsive risk-based regulation would thus ideally seek, first, to identify the regulatory logics engaged in different regulatory tasks and, second, to deal with interactions of logics – as these may combine and interplay in varying ways across the five core tasks of regulation. Risk-based regulation also has its own logic which may be incompatible with other prevailing logics, as discussed below. Thus, for the purposes of detection, regulators might use a risk-scoring system to identify certain firms as high priority risks and might deploy a number of further strategies in order to amass the information that they need to collect in order to produce risk assessments (Black 2008). They might, for instance, use a mixture of training and educational measures, co-operative discussions, advisory notes, economic incentives and legal commands in order to produce data. Being ‘really responsive’ demands an awareness of the ways in which these different strategies both combine and contribute to the broader risk-based system – of, for instance the effect that using legal commands plus criminal sanctions will have on the productivity of co-operative discussions.

In devising new responses to regulatory challenges, similar attention would be paid to compatibilities of powers and tools. Risk-based regulation, however, gives rise to numerous challenges regarding such matters. It is a strategy that requires a significant amount of information, but some regulators simply do not have powers to require information from firms (Black 2008; IOPS Report 2007). Others are restricted from adopting a risk-based approach by legal provisions which require them to inspect firms at legally specified intervals. Similar legal constraints can cut across regulators’ ability to enforce a risk-based system and assess its success or failure. Indeed, the performance criteria which are used to assess regulators themselves can inhibit them from adopting risk-based systems. It was only when the Health and Safety Executive’s (HSE’s) sponsoring department moved away from using the number of inspections carried out as a key performance indicator, for example, that the HSE could move to a more risk-based approach. A similar experience existed in the context of food safety, where European directives used to require inspections, and did not regard any other form of intervention as an adequate control (Black 2008).

In carrying out these different tasks, moreover, different strategies, powers and tools might be used in different combinations. Thus, a regulator might employ a combination of deterrence and educative strategies in order to encourage a firm to reduce risks but it might apply a set of incentives together with a selection of disclosure rules in order to assess whether regulation was reducing risks. One consequent challenge would be to analyse how the use of, say, a deterrence approach to compliance would impact on the firm’s attitude to regulation and how the risk scoring system will take account of such an effect. Another
challenge would be to calculate whether an educative strategy adds or detracts from the use of a deterrence strategy.

A further set of questions concerns the degree to which using certain strategies to discharge some tasks will impact on the use of other strategies in relation to other tasks – for example the extent to which applying deterrence–based enforcement will impact on the detection and information collection functions that are central to the risk-based regime. This can pose significant practical difficulties for risk-based regulators. Such regulators need considerable information from firms to sustain their oversight. They may, however, have to use formal enforcement actions, such as fines, to change the behaviour of many firms. In such circumstances responding to non-compliance with a deterrence approach may cut across the ability to detect that non-compliance in the first place. Firms know that any information they give to the regulator may potentially be used against them in an enforcement action and this can have a chilling effect on their cooperation with that regulator. A good, albeit anecdotal, example is the contrast in enforcement approaches of the two Australian financial regulators, APRA and the Australian Securities and Investments Commission (ASIC). APRA has a model of intensive supervision for its high risk financial institutions, but this does not involve using formal enforcement actions. ASIC, on the other hand, has moved to a much more deterrence-based approach. The consequence for their respective monitoring functions was noted recently by an Australian lawyer, who quipped, ‘When APRA asks for information, firms give it to them; when ASIC asks, they call their lawyers.’ (Note on file with author).

Finally, risk-based regulation has its own logic - that of risks and outcomes. Risk-based regulation starts with identifying risks to be managed, not rules to be complied with. The logic of risk and outcomes can cut across the logic of compliance. In a risk-based system of monitoring, officials are likely to find that non-compliance with certain rules does not in fact have an impact on the risk or outcome they are concerned with, suggesting of course that the rule is otiose, or at least there is a mismatch between the rule-makers’ perception of what is an appropriate conduct or process, and that of the supervisor, inspector or enforcer. On the other hand, an official monitoring a firm under a risk-based system may identify risks that she thinks the firm should address, but which are not covered by any rule that would provide a legal basis on which to require the firm to take action. The mismatching logics of risk and compliance can thus produce significant lacunae in the regulatory regime at the point of implementation and enforcement.

**iv) Responsiveness to regime performance**

Really responsive risk-based regulators will be performance sensitive – they will be capable of measuring whether the enforcement tools and strategies in current use are proving successful in achieving desired objectives. Such regulators will also operate systems that allow them to justify their performance to the public and other interested parties. They will also be able to adjust their strategies in order to improve on the levels of performance that they have assessed. Many of those engaged in risk-based regulation are conscious of the need to ensure responsiveness to regime performance, but finding appropriate ways to assess the outcomes of regulation can be extremely challenging. The issues of performance sensitivity and performance justification are discussed in this section and the question of performance
adjustment will be covered in the next section – which deals with adaptation to change more generally.

Performance sensitivity

Performance sensitivity will demand a programme for assessing the performance of the existing regime across the five core tasks of regulation. It will also require an understanding of those activities that detract from (or potentially detract from) the achievement of objectives but which are beyond the scope of the current regulatory regime or are ‘off the screen’ in the sense that they are going undetected. In order to set the basis for such sensitivity, the regulator must, first, be clear regarding the objectives of the regulatory regime and the link between this and the risk-based system of control. For many UK regulators, the route to such clarity involves a translation of statutory objectives into a statement of ‘risks to objectives’ and the construction of a risk-scoring system on the basis of these risks (Black 2005).

For risk-based regulators, however, performance sensitivity presents two particular difficulties. First, risk-based regulation is by its nature orientated to the future, to what may occur. If a risk does not crystallise, it can be difficult, if not impossible, to show that was the result of the regulator’s actions. Proving counter-factuals is notoriously difficult. Was the lack of an outbreak of salmonella in the last year, for example, due to the local authority’s excellent monitoring, or to improvements in food processing demanded by retailers, or to food producers’ own efforts independently of the supply chain, or simply luck? Performance assessment is not impossible, however, it is just difficult. Where a regulator is operating in an environment where there are numerous incidents occurring of a similar nature, and/or where data is easy to collect, then performance assessment is more straightforward. Environmental regulators, for example, can use data on environmental quality to assess the effect of regulation, though establishing causal links between particular strategies and environmental quality can be difficult. In the area of health and safety, there are particular types of recurrent injuries at work, such as ‘slips and trips’, which occur with such regularity that together with data obtained through firms’ incident reporting obligations, the HSE is able not only to collect data on their occurrence, but to identify from that where ‘slips and trips’ are likely to occur, and in what circumstances in different industry sectors. It can then correlate changes in their incidence with changes in its own implementation policies to derive some indication of performance, albeit an imperfect one (e.g. HSE 2005).

Where regulators are regulating situations of uncertainty, rather than ‘risk’ as such, then, by definition, there is no comparable set of data available on past incidences from which probabilities can be derived or against which regulatory strategies can be correlated. Even in these situations, some regulators are starting to perform ‘back-testing’ assessments of individual firms’ risk assessments, however. These are used to see a year later, for example, whether the risks that the regulator thought may emerge have emerged, or whether management and control systems were adequate to handle those which did emerge. As noted above, the results of such back-testing can be informative, and in that example, the regulator
adjusted the weighting of the assessments of management and control to allow for the revealed bias.

The second main difficulty that risk-based regulators have in dealing with performance sensitivity is that the risk-based approach, like many other forms of regulation, involves delegation of many regulatory functions to the firms being regulated. It is a regulatory method that focuses attention on the quality of a firm’s internal controls – in a process now commonly termed “meta-regulation” – and is observable across a range of regulatory domains, and indeed countries (Braithwaite 2000; FSA 2000: Introduction; Parker 2002; Coglianese and Nash 2001; Braithwaite, 2003; Walker 1969; Braithwaite and Drahos 2000; Black 2005). The FSA, for example, is legally required to take into account the responsibilities of senior management in performing its regulatory functions, and, to this end, it has devised a complex set of requirements for senior managers and other “approved persons” which go well beyond the normal bounds of directors’ duties both in scope and content, and for which the individual may be sanctioned for non-compliance (Black 2005: notes 185-6; Gray and Hamilton 2006).

Reliance on, and hence assessment of, a firm’s internal controls is, accordingly a central element of such risk-based frameworks – and is often seen as inevitable because regulators simply do not have the resources to do anything else. Such a layering of regulatory controls makes performance assessment particularly difficult though. A special problem is that different actors – be they corporations, regulators, credit ratings agencies or other bodies – may use different models or ‘codes’ to evaluate risks. This means that in so far as regulators enrol such actors in the regulatory regime, this involves a substituting of these codes for the judgements and decisions of the regulators – and it does so in a manner that renders risk evaluations all the more opaque to regulators as well as to the broader community. When layers of such codes are involved in the provision and evaluation of services – as when corporations, credit ratings agencies, financial institutions and regulators are involved with a securitised product – there can be a worrying lack of connection between the regulator and the risk evaluation (Gerding 2008; Gray 2009).

Apart from the difficulties of modelling risks and amassing information in a transparent and consistent manner, there are cultural divergencies that affect regulators’ abilities to assess performance by relying on firms’ systems. Regulators think of objectives with reference to statutory purposes but firms will see internal controls as properly directed at ensuring that the firm achieves the objectives it sets for itself: namely profits and market share. The risks that the regulator is concerned with will, indeed, not always be the same as the risks the firms are focussed on. Firms are ultimately concerned to ensure their own survival and growth, and if publicly listed, to maximise their returns to shareholders, despite the inroads of corporate social responsibility initiatives. The firms’ risk management systems may be adequate for achieving the firms’ own objectives but they may not be adequate for achieving the regulator’s objectives. Such divergencies mean that regulators can never rely on the firms’ own systems either for implementation or for undistorted feedback or results.

A good example of the potential divergence between managerial and regulatory systems comes, again, from financial services. This is the UK FSA’s ‘Treating Customers
Fairly’ initiative. This was an intensive, principles-based monitoring system which required firms to demonstrate that they were ‘treating customers fairly’ at every stage of the process of formulating, marketing and selling retail financial products. The FSA also required firms to put in place various mechanisms to assess their own performance. One of the strategies the firms used was a programme of customer satisfaction surveys. The FSA, however, disputed the use of consumer satisfaction as an appropriate performance metric for demonstrating compliance with the requirement that firms had sold customers products which were suitable to the financial objectives and risk appetite. The rationale for the suitability rule was that investment products are notoriously opaque and consumers have considerable difficulty understanding them. (Consumers could be satisfied with the advice received, but their limited ability to assess that advice might mean that such satisfaction was ill-founded.) Firms, however, judged their performance on the likelihood that the customer would come back again, and could see no difference between suitability and satisfaction (FSA 2008).

The further complication to be noted by the really responsive regulator is that the degree to which, and the way in which, assessment procedures can be ‘delegated down’ to firm’s internal control systems will vary across the tasks of regulation. Thus, relying on the firm’s risk management system to assess the detection of risks may involve quite different challenges to those involved in relying on such systems to assess enforcement performance. Across the tasks there may be quite different kinds of divergence between the regulators and the firms with respect to such matters as the treating of evidence as relevant and how key concepts such as ‘compliance’ are defined. For risk-based regulators, the special difficulty lies in assessing the complementarity or dissonance between the firms’ evaluation systems and the risk frameworks that reflect the regulator’s objectives. Correlating the risk evaluation systems of the regulator with those of the firms may prove hugely demanding especially if variations across firms are to be taken on board.

**Performance Justification**

As indicated above, really responsive regulators will be well-placed to justify their actions by making reference to assessments of performance and to undertakings on delivery. The difficulty that flows from espousing risk-based regulation is that this is an approach that makes lavish undertakings. Central to these is the promise that the challenges and complexities of regulation can be rationalised, systematised, ordered, managed, and controlled. As has been said of risk-based regulation: ‘…it suggests that the notoriously complex task of regulating can be rendered manageable, and that the contingencies of unpredictable events can be made controllable….hesitancies are lost in the confident exposition of risk identification, assessment and validation’ (Black 2005). The promise of risk-based regulation is thus commonly seen as the delivery of a system that is not only more rational, cost-effective and controllable than other systems but more transparent and more easily deployed as a means of justifying regulatory actions and policies (Better Regulation Task Force 1999, 2002, 2004a, 2004b; Hampton 2004; Gershon 2004: para.2.22; O’Donnell 2004: para.3.80; Hood et al 1999). At the political level, risk-based regulation was welcomed
in the UK as a way to curb what was seen as the insatiable appetite of politicians and the wider public for regulation (Black 2005).

Delivery on these undertakings is extremely difficult, not least because there is often considerable dissonance between the regulator’s understanding of risk priorities and those of the firms, or indeed the wider public. Choosing which risks to focus on is a political, not a technical issue and judgements have to be made on such matters as: whether to target the largest risks or the places where the largest risk reductions can be effected for a given level of resource input; whether to focus on individual risk creators or specific types of risk; the right balance between acting on systemic risks and controlling individual risks; and ultimately what is an acceptable level of risk. A further issue of difficulty is whether to err on the side of over intervention (assuming that certain firms pose risks when they do not) or of under-intervention (assuming that they do not pose risks when they do). Each position on these issues brings contention and problems. One danger, for instance, of focusing only on firm-specific risks is that regulators may pay too little attention to the potentially huge cumulative effect of particular types of compliance failures across firms (Black 2005: notes 141-151). Another danger, as noted above, is that the regulators may also fail to pick up on trends of events that are external to the firms and which might be relevant to their risk assessments.

Nor does transparency always assist in legitimation. Regulating according to a risk-based framework exposes the reality that there will be a limit on the resources that can be spent on controlling certain types of risk creators (e.g. low impact firms), or on firms in certain cases (for example, medium/high impact but low risk). Such a framework also exposes the balance between individual and systemic risk controls. This means that officials are required to leave certain risks or types of risk uncontrolled or subject to limited supervision. This can be difficult for regulatory managers to justify to the public and to regulatees. After a harmful event has occurred at a firm it may, for example, be difficult to explain to the media that firms of that class are not regulated as a priority because they have risk scores that are too low. Similarly, it may be difficult, after a systemic catastrophe such as the credit crisis, to explain why systemic issues had not been given a higher priority.

Risk-based systems may also be difficult to justify to staff who have to reconcile the advertised technical neatness of those systems with the political messiness of the world that they seek to control. An irony here is that risk-based regulation was sold within some regulatory bodies as a defence mechanism – on the basis that if one followed the book, one would be immune from criticism (Black 2005: notes 172-3). The bureaucratic reality, however, has been that if senior management fail fully to articulate the extent to which they will “buy in” to the risk-based process so that they accept that mistakes will be made and that things will be left undone, this will reduce the confidence of staff in the system and lead to the taking of self-protective steps such as operating on the basis of factors other than risk analyses (such as perceptions of political risks to themselves). This is liable to hamper the implementation of the risk-based regime and reduce the rationality that lies at the heart of its justificatory claims.

A special justificatory challenge for risk-based regulation is to satisfy expectations regarding openness, transparency and accountability. Here again there are paradoxical elements. On the one hand, risk-based regulation holds out the prospect of transparency
through the exposure of its numerical/analytic basis. On the other, a closer look at the operation of such systems reveals that they are not Richard Rogers edifices with the works on the outside. Not only are they built on high levels of discretion and politically contentious judgements but the important policies and decisions tend to be hidden away behind the apparently neutral language of the risk assessment model (Black 2005). This point applies especially to the definitions of thresholds for intervention, the risk assessments themselves, and the subsequent categorisations and scoring of firms.

All regulators have to prioritise the use of resources but the message of a really responsive approach is that it is best to understand the implications of doing this through risk-based processes. What can be seen is that framing the regulatory task in terms of risk involves buying into particular conceptions of the problem at hand, leads to the framing of a solution in a particular way and produces special challenges of justification and legitimation. In adopting risk-based frameworks, regulators attempt to define the acceptable limits of their responsibility and hence accountability. It is also the case that matters of internal operation, notably the construction of the risk-scoring system, are where the real choices are being made about what matters to that regulatory agency and what does not: in particular about the selection of risks or levels of risk that the regulator is, or is not, going to take action on. In such processes risk-based regulation implicitly or explicitly defines what risks the regulator should be expected to prevent, and those which it should not - those it should be blamed for not preventing and those which it should not be blamed for not preventing. As Power has argued, attempting to manage risk requires a new ‘politics of uncertainty’, involving assessments not only of who should be responsible for dealing with its consequences but an appreciation that no-one is to blame for true uncertainty (Power 2005). Risk-based regulation also requires a new and related politics of accountability and a quite distinct mode of legitimation. That politics involves new debates on who should be making decisions on the risks that are important and those which are not.

Such issues of justification and legitimation, moreover, can be expected to vary across the core tasks of regulation. The above discussion of risk-scoring, for instance relates to the task of discovery – of identifying targets for intervention. Quite different, but nevertheless risk-based-specific sets of issues will arise when looking at the processes of enforcement and compliance-seeking on the ground, or the processes of assessing performance by means of risk analyses, or those of modifying regulatory strategy. The need to come to grips with all of these issues is the important message of the really responsive approach.

v) Responsiveness to change

A really responsive framework suggests that regulators and regulatory systems should be dynamic, and capable of building on their performance sensitivity by learning from and improving on their past performance. It also suggests that regulators should be able to adjust to the changes that constantly impact on regulation – such as shifts in objectives, the advent of new risks and the emergence of new risk creators. These capacities to cope with change
should, moreover, be reflected across all of the core regulatory tasks. Responsiveness, to change, in addition, should involve a willingness and capacity to re-think regulatory strategies quite radically and to contemplate completely new forms of control mechanism if performance assessments indicate that these are necessary. For risk-based regulators, however, such responsiveness involves quite particular challenges.

With regard to detection work, a central such challenge is the uncovering of new risks and risk-creators. In a risk-based regime, the inherent danger is that of ‘model myopia’ – that regulatory officials become committed to an historically-captured set of risk indicators and assessment criteria. Such commitment inhibits the regulator from responding to an unpredicted future. If the safest thing to do is to follow the risk framework, the way of least resistance is not to respond to any circumstances or events which are not anticipated by that framework. The irony is that risk-based frameworks are in danger of becoming institutionalised in a way that negates their capacity to deal with the very predictive inability that they are intended to meet.

The credit crisis also showed that a further aspect of this danger of institutionalisation is that regulators may become committed to a risk model that is technically or intellectually deficient in a manner that prevents adaptation to developing threats. Thus, an error of regulators in the period leading up to the credit crisis was the assumption that the securitisation process led to greater risk spreading between institutions and individuals, and across borders, and so produced safer, more risk resilient, as well as more efficient financial systems. The shocking reality that the risk-based regulators had to come to terms with was that they had failed to anticipate the way that institutional and trading interconnectivities within the global markets tended to lead to increasingly dangerous concentrations of risk and to higher, linked, vulnerabilities across the board. These developments eventually produced a collapse in financial confidence that the regulators could not keep at bay (HM Treasury 2009: para. 3.30; US Government Accountability Office 2009).

The Hampton Review, indeed, endorsed the general strategy of risk-based regulation but it was well aware of the danger that risk frameworks can prove too static. Hampton argued that regulation ‘should always include a small element of random inspection’ in order to check on the validity of the risk assessment system (Hampton 2005: para 2.38; Statutory Code of Practice for Regulators 2007: para. 6.2). A value of random inspection, on this view, is that it holds out the prospect of uncovering new risks and risk-creators in a way that is unlikely to flow from an inspection programme that is based on analyses of previously-identified risks.

Random inspections may also act as a useful deterrent to non-compliance, but as techniques for combatting model myopia they are only consistent with a really responsive approach if the random inspections are performed in a way that breaks away from the normal risk assessment framework. Only then will they pick up risks which may be systemic, or at least which are not unique to that particular firm. Here again the institutional environment of the risk-based system becomes important. Regulatory supervisors or inspectors need to feel able to communicate new or emerging risks to those in the policy or risk division in the course of their day to day monitoring and assessment activities. At best, there should be an institutionalisation of a dynamic process for identifying new or emerging risks. Some
financial regulators, such as Canada's OSFI, initiated such an process during the financial crisis, forming an ‘emerging risks committee’ comprised of supervisors and senior management to systematise the identification and assessment of rapidly emerging risks, and analyse their impact on the banks they regulated.

A second need for the really responsive risk-regulator is to react to change by developing new rules and tools that will assist in detecting undesirable risk creation and in producing compliance with relevant requirements. The institutional environment may inhibit this, however, as the regulator may not have rule making powers, but has to rely on a legislature, or in the EU context, the EU law making institutions to change the legal requirements. Dynamism and reflexive responsiveness is hard to achieve in these situations.

A third need is closely related and that is to ensure that enforcement and compliance-seeking activities deal successfully with new risks and new risk-creators. On these fronts, however, there are problems that flow directly from adherence to risk-based regulation. One problem is that if the regulators see the world through the lens of a given framework for risk analysis (and delegates much of the application of this to firms), it is very difficult, as noted above, for them to come to grips with new risks and risk creators by developing new control responses. It may be especially difficult to move beyond the envelope of a risk framework if those new risks and risk creators develop in another arena or on a different scale (e.g. global rather that national) – as in the credit crisis of 2007-9 when national regulators were slow to respond to developing problems of global system-wide risk (HM Treasury 2009).

In the case of new rules and tools to assist in compliance-seeking, the additional difficulty is that a risk-based system may, as already indicated, focus more on identifying risks than on ways to secure compliance or reduce those risks on the ground. The devising of new, adaptive, rules will, accordingly have to be guided by approaches and theories – such as compliance, deterrence, or ‘escalating response’ strategies - that go beyond the remit of most risk analyses. Here, though, there is a further problem. It is no simple matter to prescribe that, for the purposes of compliance-seeking, risk-based systems should be combined with other systems. In the first instance, officials who are committed to a risk-based system (or whose managers are so committed) may prove unresponsive to changes (either in legal powers or approaches to compliance-seeking) because they are reluctant to move away from a ‘pure’ risk-based system. Second, where mixtures of risk-based and other approaches are used, these mixtures cannot be assumed to be free from tensions – a point suggested in discussing regulatory logics above. Where, for example a regulatory official allocates enforcement resources to deterrent activities in order to further compliance, this may involve prioritising the use of resources in a manner that departs from the prescriptions that would flow from the relevant risk analyses. A similar effect is apparent when the regulator diverts attention to investigation activities prior to bringing enforcement actions. The HSE, for example, spends about 60% of its inspection resources on investigations after accidents prior to bringing prosecutions. One reason for this is that firms now are more likely to contest prosecutions more vigorously, as a conviction for health and safety violation prejudices their applications for public procurement contracts. This is a good example of the impact of the broader institutional environment on the operation of a regulator’s risk based approach. The risk based system, as a result, loses its purity and the risk analysis loses its power as a driver of
the regime - which now operates as a hybrid risk-based/deterrence/incident-reactive system. The same kinds of ‘corruption’ of the risk-based regime may occur where reference is made to other modes of compliance-seeking and would be likely to involve either or both resistance from those committed to the risk-based philosophy and a confusing hybridity.

The fourth and fifth core tasks of regulation involve the assessment and modification of performance. Risk-based regulation might be expected to cope well with these. After all, the risk-based approach - at least the ‘technical’ vision of this – offers a ready means of judging performance. The risk scores of regulated firms and individuals can be compared year on year (or month on month) and this will reveal whether overall levels of risk are increasing or decreasing. This can provide some value, but as noted, more sophisticated systems of back-testing can be more useful. Simply comparing risk scores year on year is prey to ‘gaming’ by officials, and even in its absence, it is limited as a method of performance assessment as it only tests the model within its existing parameters. This poses a number of difficulties. First, such comparisons tend to focus on a given set of historically-established risks and, if this is so, they will reveal little about the regulator’s success or failure in coming to grips with new risks and new risk creators. Second, a given framework for risk analysis will presuppose that there is a perfect fit between the risk framework and the regulator’s objectives – and it will accordingly give no indication of the extent to which undesirable risk creation is escaping the regulatory net. There is liable to be no measure, for instance, of the prevalence of creative compliance or new types of risk creation or risk creators that are not covered by the current rules. As a result, analyses of relative risk scores will not indicate whether the regulatory regime is addressing a major portion of the challenges faced or only a small percentage of these.

A third problem that risk-based systems may encounter in dealing with change is that such systems may make it difficult for regulators to adapt to shifts in preferences and objectives. Risk-based regimes always have to contend with possible disjunctures between the regulator’s perceptions of risk and those of the public (or certain groups of interests) but the additional problem to be noted here is that those disjunctures are not static. Preferences concerning regulation often change – as seen in the post credit crisis period in the UK when sections of the public, the Government, the regulators and the media lost a good deal of faith in ‘light touch’ regulation. The public may want different things of regulators at different times and so may governments, legislators, extra jurisdictional bodies and particular groups of interests. The problem is that risk frameworks tend to lock regulators into a certain mode of analysing risks, a certain position on identifying which risks are important and a particular position on the thresholds for acting to control risks. If regulators are committed to the framework, they may be slow to respond to changes – especially when the processes of constructing and developing that framework are positioned deep within the bureaucratic process and are, as a result, insulated from the public pressures that might be expected to galvanise change.

A further problem is that if risk-based regulation involves a misalignment between the institutional risks of the regulator (i.e. the risks to the regulator’s reputation and objectives) and risks to society (of harms such as injuries or deaths), this may make the regulator unresponsive to changes in risks to society’s interests. Thus, to cite Rothstein et al’s
example of rail safety (Rothstein et al 2006), a rail regulator may tend to focus on risks of multiple fatality accidents to a degree that is not commensurate with the attention it devotes to the control of common minor accidents that cumulatively cause as many or more fatalities. It may do so because it is aware that multiple fatality accidents may detract from the regulator’s reputation in an especially negative and disproportionate manner. Controls of risks to the regulatory institution and risks to society are, for this reason, disjointed – but disjointed in a manner that may make the regulator less responsive to changes in risks to society (e.g. new kinds of minor accident) than it might be if it operated a regulatory system other than one that gives risks to the regulator’s reputation and objectives a central place.

Some regulators manage these challenges by incorporating public perceptions explicitly into their risk-based frameworks. For the UK Pensions Regulator, for example, a key criterion for assessing risk is whether a failure in a particular area would lead to a loss of public confidence in the regulator and in the pensions system. Others manage it on a more ‘fire-fighting’ basis. As the political salience of an issue increases, the regulator’s risk tolerance decreases, or as one financial regulator commented during the crisis, ‘events drive you up the probability curve’ (Black 2008). It may be argued that such responsiveness to political salience is appropriate, and a good example of ‘spontaneous accountability’: the regulator making itself accountable to political demands. However, if a regulator is always responsive to political assessments of risk, as amplified by the media and others, then risk based regulation loses much of its identity as a systematised and rational way for regulators to manage their resources.

Finally, note should be taken of the special difficulties of contemplating radical changes in regulatory strategy from ‘within’ a risk-based regulatory regime. The problem here is that a mindset that centres on analysing and reacting to risks will not be readily attuned to the consideration of ways in which risks can be ‘designed out’ of economic or social processes by moving towards pre-emptive managerial strategies. Such shifts of approach may demand a breadth of analysis that the ‘process myopia’ of the risk-based system does not encourage. The message of the ‘really responsive’ approach here is that the risk-based regulator should always be aware of the possible need to move beyond the merely responsive.

Conclusions

A ‘really responsive’ approach to risk regulation brings two central messages. The first is that it is best to regulate in a way that is responsive to: regulated firms’ behaviour, attitude and culture; institutional environments; interactions of controls; regulatory performance; and change. The second is that the challenges of regulation vary across the different regulatory tasks and that astute regulators will deal with the variety of those challenges. We have argued here that risk-based regulation has achieved broad acceptance within many governments and regulatory organisations but that there is a need to think in a more structured manner about the ways in which risk-based regulation can come to terms with the many hurdles to be overcome if it is to succeed on the ground. The ‘really responsive’ framework, we suggest, offers a basis for such structured thinking. As shown above, an application of the really responsive approach quickly exposes the kinds of question to which risk-based regulators have to find answers – but it also reveals that risk-based
regulation imposes quite distinct structures on the challenges that have to be risen to if regulation is to succeed. In addition, the really responsive approach forces us to shift our conception of risk-based regulation so that it is seen not as a free-standing and technical guide to regulatory intervention but as a particular way to construct the regulatory agenda and as a control strategy that has to be combined with other control strategies in different (and often contentious) ways across different contexts and regulatory tasks.

Being really responsive involves coming to grips with numbers of issues in a manner that might appear to be daunting and difficult to operationalise. Those difficulties are, however, vastly outweighed by the costs of regulatory failure. If there is one message to take from the credit crisis, it is that there is a colossal price to pay if regulators do not deal adequately with the challenges discussed here – notably, in the case of banking, those produced by changes in the nature of risks or risk creators and by the constraints that flow from the institutional environments in which the regulators and regulated firms work.

NOTES
1. ‘Risk-based regulation’ has a range of meanings- see Black 2008.
2. The evidence is mixed regarding the culpabilities of risk-based regulation, as banks in Australia and Canada, whose regulators have well-developed systems of risk-based regulation, fared far better than those in other Western countries, suggesting that the causes of regulatory failure were more complex than are accounted for by the existence of a risk-based system of supervision.
5. Verantwoordelijk voor wonen, ruimte en milieu (Ministry of Housing, Spatial Planning and the Environment).
6. In the UK the Statutory Code of Practice for Regulators 2007: para. 6.3 states that regulators ‘should focus their greatest inspection effort’ on regulated entities where a breach would pose a serious risk to regulatory outcome and where there is a high likelihood of non-compliance by regulated entities.
7. There is a danger in compounding quantitative risk analyses with quantitative ‘amenability analyses’ so that a kind of ‘quantitative ritualism’ results. Similar scepticism applies to seeing amenability issues in 'risk' terms - for a discussion of the diminishing returns that can be achieved by seeing all issues as 'risk issues' see Rothstein et. al. (2006).
8. For example, meat hygiene regulation; and until recently, environmental regulation: EC 882/2004, Official Controls performed to ensure the verification of compliance with food and feed law, animal health and animal welfare rules.
9. Sparrow would refer to these as ‘invisible’ offences (Sparrow 2000:192, 272-3). See also the discussion of the FSA’s ‘watch lists’ in Black 2005: note 111.
10. See, for example the FSA’s *Reasonable Expectations*, which noted the gap between public expectations, and what “reasonable” expectations should be. The paper made it clear that “non-zero failure” meant that the regulator would not, and should not be expected to, prevent every “negative event”: every financial failure of a firm, every incidence of non-compliance, every incidence of market failure.

**REFERENCES**


Haldane 2009; Gerding 2009