RETHINKING RISK

A PILOT MULTI-CRITERIA MAPPING OF A GENETICALLY MODIFIED CROP IN AGRICULTURAL SYSTEMS IN THE UK

A report by SPRU in association with Unilever,
and with help from Green Alliance
and a variety of other environmental and consumer
non-governmental organisations.

Andy Stirling
Fellow
SPRU science and technology policy research
University of Sussex

Sue Mayer
Director
GeneWatch UK

CONTENTS

1. EXECUTIVE SUMMARY

Introduction

There is an urgent need for new ways to manage technological risks. Climate change, the ozone hole, nuclear waste, pesticides, hormone disrupting chemicals, BSE, Brent Spar, genetically modified food - a host of apparently intractable risks pass in and out of the headlines at a frenetic pace. The issues quickly become polarised. There are signs that public anxieties over each successive 'revelation' of technology-induced threat are compounding into a corrosive general attitude of fatalism, disillusion and distrust. Reassurances on the part of government or industry are increasingly coming to be seen as cynical exercises in financial or political damage limitation. The established techniques of risk assessment seem unable to accommodate the wide diversity of issues or address the powerful emerging forces.

In looking for new approaches, risk appraisal tools are required which are: flexible and broad in scope; open to divergent interests and values; able to acknowledge uncertainty; whilst being systematic, transparent, verifiable and accessible as well as practically feasible and efficient. Conventional risk assessment methods fail to meet many of these needs. In part, this is because they seek to provide a single (apparently uniquely authoritative 'sound scientific') answer: *safe* or *unsafe*. Yet it is increasingly recognised by bodies such as the US National Research Council[1] and Presidential Commission on Risk Assessment[2] and the UK's Royal Society[3] and Royal Commission on Environment and Pollution[4] that attempts to assert uniquely 'scientifically sound', 'economically rational', or otherwise ostensibly definitive conclusions can all-too-easily leave crucial subjective factors concealed.

It is ironic that efforts to mould consensus through the assertion of a particular set of values in risk assessment can actually foster greater controversy and mistrust. What is lacking is an approach which provides a reliable 'map' of the key technical and social issues from which policy makers might hope to make more transparent, robust and readily accountable decisions. Multi-criteria appraisal techniques - often used in planning and technology assessment in some countries, but only rarely in response to contemporary risk issues in the UK - provide one potential approach.

In this pilot study, some of the techniques of multi-criteria appraisal were applied to the comparative evaluation of the usefulness of genetically modified (GM) crops – an issue which is subject to widely divergent views. The approach is called 'multi-criteria mapping' (MCM) because the intention is to demonstrate how a risk debate can be 'mapped' – establishing the main contours and identifying the key areas of difference and convergence. The particular case examined was that of GM herbicide tolerance in oilseed rape, with a comparison made with other strategies for the cultivation of oilseed rape.

Methodology

The pilot MCM process is described in Box 1. It involved twelve participants, each prominent in the current debate over GM crops and chosen to reflect a wide range of institutional interests and perspectives (from strongly favourable to strongly opposed to GM strategies). They included regulators (2), academic scientists (2), representatives of biotechnology industry and the food supply chain (4) and a variety of religious and public interest groups (4).

The features of the MCM approach which distinguish it from a conventional risk assessment include:

- A focus on comparing the *relative* performance of a range of different options (here including organic, conventional and GM strategies for the production of oilseed rape), rather than simply on asking whether a particular individual option is 'safe' or not when taken in isolation.

- The criteria used to evaluate the options were chosen by the participants themselves, extending beyond simple quantitative factors and with rela-

1 NRC, 1996
2 EPA, 1997
3 Royal Society, 1992 Chapter 5.
4 RCEP, 1998

BOX 1:

Multi-criteria mapping of a genetically modified crop in agricultural systems in the UK -Methodology

To consider the *relative* performance of genetically modified, herbicide tolerant, oilseed rape, six basic policy options were compared (plus up to six others chosen by each participant):

- No GM crop, organic agricultural system
- No GM crop, integrated pest management system
- No GM crop, conventional agricultural system
- GM crop with segregation and labelling
- GM crop with post-release monitoring
- GM crop with voluntary controls on areas of cultivation

Step 1: Individual interviews (2-3 hour session) in which the participant:

- selected any additional options they wished to consider
- defined the criteria that they considered should be used to evaluate the options
- scored the options under each criterion, specifying uncertainty when relevant
- decided upon the relative importance weighting for each criterion

The weighted sums of the scores under all criteria for each option were calculated on a simple computer programme to give a picture of overall performance. The options were then ranked accordingly.

Step 2: Analysis (Qualitative and Quantitative) To compare and explore differences between individual participants, the researchers then:

- grouped criteria into 5 categories: environmental, health, agriculture, social and economic.
- identified areas of technical agreement/ disagreement
- examined how uncertainty affected the outcome patterns
- conducted sensitivity analysis to determine what was driving the outcome
- investigated the effect of considering a diversity of options

Step 3: Feedback on preliminary results and re-evaluation

Participants considered their own results in comparison with that of other participants and were asked to reassess or confirm their initial input. The results and analysis were adjusted where necessary.

tively *few constraints* imposed by the technique itself.

- The 'scoring' of options under the different criteria provided a *multidimensional* picture of performance, rather than being based on a single absolute yardstick.
- The full scope of technical and scientific *uncertainties* was highlighted by explicit attention to 'pessimistic' and 'optimistic' assumptions in scoring. This contrasts with a focus on those single values deemed to be most 'likely' or 'appropriate'.
- Different judgements over the relative importance of different criteria - something which is often concealed in risk assessment - were here accommodated and explored by the use of explicit numerical *weightings*.
- As well as permitting conclusions concerning overall performance, the final picture of the relative rankings of the different options provide a clear reflection of how the options perform differently *under different perspectives*.
- The techniques of *sensitivity analysis* were used quite intensively to investigate the key determinants of the resulting picture of performance.

For all twelve participants the procedure yielded a wealth of information concerning the particular options which appeal to different constituencies, the different types of appraisal criteria and the ways these are framed under different perspectives. However, two of the participants did not complete all parts of the process because they felt unable to assign numerical scores or weightings for their evaluation criteria.

Findings and Conclusions

The MCM method successfully engaged a wide spectrum of interests in a manner which is often difficult in such hotly disputed controversies. As a result, the range of perspectives accommodated in this study was much wider than that which is embodied in conventional approaches to the risk assessment of GM crops. The result was a fairly comprehensive 'map' of the issues surrounding the use of GM oilseed rape, providing a relatively full and detailed reflection of the key themes currently raised in public debate.

A very wide range of criteria (117 in total) were defined by the different participants, covering environmental, agricultural, human health, social and ethical issues. Likewise, a wide variety of different options were identified and their implications explored. For no participant were all their criteria included in the existing formal system for the regulatory appraisal of GM crops in the UK. Significantly, broader non-technical issues such as ethics and institutional demands were considered to be as relevant to the appraisal of environment and health effects as more technical matters such as gene flow and toxicity.

Although there were stark contrasts between the final rankings obtained by the different participants, there were also some common themes. Across all perspectives, the organic option performed relatively well, not only under environmental criteria (where it emerged unequivocally well), but also in relative terms under all criteria taken together. There was also a general picture of the relatively poor performance of conventional intensive agriculture. GM options were found to perform best only under the perspectives of certain government and industry participants. The degree to which these findings accurately reflect some currently emerging trends in the debate suggests that a MCM approach may have the potential to provide decision makers with reliable - and correspondingly valuable - insights into crucial risk issues and their policy implications.

The outcome of the present MCM was largely determined by the type of criteria that were chosen (by the particular individuals involved), the way these were defined and prioritised and the assumptions employed in scoring (the so-called 'framing assumptions'). Perhaps surprisingly, however, the values taken by the criteria 'weightings' evidently had relatively less impact on the final picture than did more qualitative factors in the characterisation of the criteria and the 'framing' of the scoring. From this it follows that the consideration of cultural, political and ethical issues separately from - or after - the conduct of risk assessment is unlikely to compensate for any constraints or idiosyncrasies in the framing of the risk assessment itself. An MCM approach may therefore assist in the crucial business of 'risk characterisation', prior to - and subsuming - the conduct of other risk assessment techniques.

An MCM approach may also help by establishing a broader and more robust basis than conventional risk assessment for judgements over what counts as 'harm'. At present, the regulatory appraisal of GM crops is conducted on the basis of comparison with the risks associated with current agricultural practices. However, the generally favourable picture of organic agriculture and the relatively poor performance of conventional agriculture under different perspectives in this exercise, suggests that progressive standards may be a more appropriate yardstick of harm than the *status quo*.

One means to implement the much-discussed 'precautionary approach' is to avoid putting all the eggs in one basket – pursuing in parallel several of the better-performing options. This study revealed support from all sides of the debate for this general principle of diversity. However, in addition to identifying convergences of views, this focus on diversity also highlights practical difficulties. For example, GM and organic farming strategies are widely seen to interfere with each other and so appear to be mutually inconsistent. Where the effect of certain options is to compromise or inhibit the pursuit of other strategies, then, recognition of the benefits of diversity might lead to such options being regarded unfavourably in appraisal.

In conclusion, the MCM process allowed constituencies with starkly divergent interests and values fully to engage in the appraisal process. This was possible because MCM provides an approach which is at the same time relatively pluralistic, systematic and transparent. Indeed, for these reasons it may be that MCM offers a means to help avoid the dangers of (on the one hand) spurious attempts to impose 'consensus' and (on the other) opaque, distrustful and corrosive social conflict over technological risks.

The present pilot study could be extended by allowing for more discussion between participants, by including different sectors of the general public in the definition and weighting of criteria and by bringing in further expertise in the conduct of the technical scoring. It is particularly important that any subsequent exercises include wider publics both to identify any contrasts with the specialist arena and to confirm and enrich the 'map' of the overall debate over the use of genetically modified crops in agriculture.

Acknowledgements

The authors are indebted to the many people who made this project possible. First and foremost, we owe a debt to the twelve anonymous participants who so generously gave their time from busy schedules during a hectic period. Very useful discussions were held with Robin Grove-White, Jane Hunt, Phil MacNaghten and Peter Simmonds at the Centre for the Study of Environmental Change (CSEC) at Lancaster University, and Erik Millstone, Paddy van Zwanenberg and other colleagues at SPRU. Thoughtful and helpful peer review comments were provided by Jacquie Burgess (ESRU, UCL), Silvio Funtowicz (JRC Ispra), Bruna de Marchi (Gorizia), Peter Nijkamp (Free University Amsterdam), Jerry Ravetz (Research Methods Consultancy), Tom Wakeford (UEL), Brian Wynne (CSEC, Lancaster), Ahti Salo (Helsinki) and Geraldine Schofield

(Unilever). Colleagues on the project Steering Group and the UK Panel on Genetic Modification gave additional useful feedback, especially Christine Drury at Unilever and Dorothy MacKenzie at Dragon International. The development of the multi-criteria mapping technique by Andy Stirling was supported as part of a fellowship under the ESRC's Global Environmental Change Programme. Responsibility for the development, implementation and reporting of this project remains entirely with the authors.

2. INTRODUCTION

2.1 The New Politics of Risk

According to an influential body of thought in the social sciences, modern industrial civilisation at the end of the Twentieth Century has seen the advent of the 'Risk Society'. Under this view, the concept of risk has become a dominant ordering principle, helping to structure and condition social and institutional relations and, to some extent, replacing monetary wealth and cultural privilege as the focus of distributional tensions and political conflict.[5] Divergent values and interests together with issues of trust, rights and legitimacy in the regulation of risk are beginning to assume at least as much importance as the more traditional scientific and technical connotations.

Despite the social complexities, the physical implications of technological risk are almost too obvious to spell out: climate change, the ozone hole, urban smog, nuclear waste and proliferation, pesticides, hormone disrupting chemicals, BSE, Brent Spar, genetically modified food. A host of intractable risks clamour for attention, threading their way in and out of the headlines at a frenetic pace. The issues quickly become polarised. There are signs that public anxieties over each successive 'revelation' of technology-induced threat are compounding into a corrosive general attitude of fatalism, disillusion and distrust. Reassurances on the part of government or industry are increasingly coming to be seen as cynical exercises in financial or political damage limitation.

With mounting institutional and economic commitments to global technological infrastructures, the stakes are high and growing ever higher. Innovation proceeds at an unremitting pace. Once a particular industrial strategy or technological path has been chosen, a host of self-reinforcing mechanisms come into play. The enormous investments of human resources, financial capital and institutional reputation can render technological trajectories once taken effectively irreversible. The world-wide experience of nuclear power illustrates the enormous costs to all concerned of (depending on your view) over-ambitious expectations, belated critical questioning or a premature 'loss of nerve' on the part of society as a whole. On the other hand, a failure to seize the initiative and harness the positive creative potential of science and technology can lead to economic stagnation and even defeat in the face of the many challenging problems of the modern world.

The question is: what road to take? Whether they result from technological hubris or a post-modern crisis of confidence, mistakes are not easy to recognise. They are even more difficult to correct. There is agreement on all sides of the debate that the more profound and pervasive long term dimensions of technological risk cannot be left to 'the market' alone to resolve. Private enterprise and public interest groups alike seek consistency, clarity and decisiveness on the part of government (and, increasingly, inter-governmental) regulatory institutions. But how are such qualities to be achieved amidst the messy and intractable complexities and uncertainties of the emerging 'Risk Society'?

It is against this daunting background that the practical business of risk assessment and technology appraisal must now take place in sectors extending across energy, chemicals, transport, information, communication and of course, food production. In all these areas, it is increasingly coming to be realised that the old 'scientistic', expert-centred approaches are not enough. The way risks are characterised, prioritised and distributed, their ethical and cultural implications and the way they are communicated and understood, collectively serve to transform an apparently narrow clinical notion of 'risk' into an intrinsically subjective and value-laden concept. The traditional 'expert institutions' of risk assessment such as government advisory committees, professional institutions and scientific associations wield undoubted specialist knowledge in their respective circumscribed fields.

5 Eg: Giddens, 1990; Luhmann, 1991; Beck, 1992; Lash, Szerszynski and Wynne, 1996.

Yet they remain no better equipped (or mandated) to decide upon profound general questions of values and interests than are any other assemblages of citizens.

Appreciation of this picture does not imply some sort of 'post-modern' rejection of the value of science. It is clear that science in the broadest sense remains an essential element in the effective management of technological risk. However, there are profound *scientific* reasons for acknowledging that science, on it's own, is never sufficient in itself to determine social decision-making on risk.[6] Where the probabilities and magnitudes associated with certain risks are themselves subject to uncertainty, for instance, it is hardy scientific to pretend that these quantities are certain. Indeed, the formal definition of the concept of risk itself implies the equally well-founded but less frequently discussed concept of ignorance – a condition under which not only the probabilities, but also the possibilities themselves are unknown.[7] It is hardly scientific to deny the possibility of surprise. Judgements over how much "we don't know what we don't know" remain fundamentally subjective.

Likewise, the appraisal of technological risks always requires the mixing together of different issues such as: fatalities, injuries and various forms of human illness, ecological, biodiversity and animal welfare issues, frequent, routine and rare catastrophic events, reversible and irreversible effects, occupational, voluntary and involuntary exposures, risks to disadvantaged groups, children and people who gain no benefit from the activity in question.[8] Here again, it is a matter of the science underlying risk assessment itself that there can be no single definitive way of combining such 'apples and pears' in a plural society.[9] Nobel Prize-winning work in economics demonstrates the theoretical impossibility of aggregating different social preferences.[10] As a result of these and other *scientific* insights, it is increasingly recognised that even the most ostensibly technical citadels of risk analysis (the assigning of probabilities and the quantitative measurement of harm) remain, at the core, fundamentally context-dependent, subjective and thence ultimately political in character.[11] In the face of these complexities, the advocacy of 'sound science' as the *sole* determinant of decision making on risk amounts to little more than *scientistic* rhetoric. It is ironic that the aspiration to an exclusive dependence on 'sound science' in risk assessment is itself profoundly unscientific.

The challenges posed by these gradually emerging realisations are now being taken up in a number of countries. Although assertions of 'sound science' and accusations of 'public irrationality' remain a feature of wider debate, authoritative expert reports now only rarely assert the exclusive sufficiency of science or frame the issues as a problem of rationality. The traditionally narrowly technical procedures of risk assessment are being complemented by greater attention to the understanding of social and ethical issues. Real efforts are being made in many areas to open up the procedures of regulatory appraisal to a wider range of constituencies and perspectives and to make the process more transparent.[12] The inclusion of local residents, lay citizens and divergent interest groups is increasingly recognised to confer greater analytical breadth and robustness as well as enhanced legitimacy.[13] Experiments are underway with novel 'deliberative' and 'participatory' appraisal procedures, such as consensus conferences, focus groups and citizen's juries. In some countries (such as Denmark and the Netherlands) trends have progressed to the point where such approaches have become a statutory part of the regulatory process. Despite important recent initiatives.[14] far greater experience has been gained in some other countries (such as Germany and parts of the US) than in the UK.

Of course, the new 'deliberative' and 'participatory' approaches are not without their own problems – not all of which are shared by traditional analytical techniques. There are as many points of tension with orthodox risk assessment as there are potential synergies. It is far from being a foregone conclusion that such approaches offer a way out of the current risk impasse. In this light, the crucial task is to evaluate the strengths and weaknesses of the new approaches in comparison with the old. There is a need for open minds, creative thinking, free experimentation and good communication across the old boundaries. It is in this spirit that the present project has been undertaken and against which the results might be judged.

6 These issues are discussed in more detail in a recent report for the European Commission Forward Studies Unit, Stirling, 1999.
7 Stirling 1998.
8 These factors are examined in the concrete case of energy technology risks in Stirling, 1997.
9 These issues are discussed in more detail in Stirling, 1997, 1998, 1999.
10 Prominent amongst this work is the infamous 'Arrow Impossibility' (Arrow, 1963), now itself the subject of an extensive literature (cf: Kelly, 1978, MacKay, 1980 and Bonner, 1986 with a convenient summary of the discussion provided in Pearce and Nash, 1981).
11 Acknowledged, significantly, in Department of Environment, 1995 paragraph 4.10 and elaborated in NRC, 1996.
12 Eg: Webler, et al, 1995; Durant and Joss, 1995; Renn et al, 1996.
13 Eg: Fiorino, 1989; Sclove, 1995; Bohmann, 1996.
14 Eg: UK National Consensus Conference on Plant Biotechnology, 1994; MacNaghten, et al, 1995; Grove-White, et al, 1997; Citizen's Foresight, 1999; Clark et al, 1998; UK National Consensus Conference on Radioactive Waste, 1999.

2.2 Multi-Criteria Mapping in a 'Market of Methods'

A wide variety of techniques have been used in the appraisal of technological risk. The more important general analytical approaches include decision and policy analysis,[15] life cycle analysis and environmental impact assessment,[16] multi-attribute utility theory and multi-criteria evaluation,[17] probabilistic, comparative and environmental risk assessment,[18] orthodox and 'constructive' technology assessment,[19] as well as the various forms of environmental cost-benefit and cost-effectiveness analysis.[20] To these might be added the various formal social scientific approaches to the structuring of qualitative deliberation, including consensus conferences, citizen's juries and focus groups.[21]

Each of these approaches displays different strengths and weaknesses. Each has its own emerging negotiated niches for application. In practice, of course, the picture will often be as much a reflection of the quality of a particular study as of the methodology in general. However, both the more traditional 'scientistic' approaches and the new 'participatory' and 'deliberative' approaches present potential difficulties. Probabilistic risk assessment and cost-benefit analysis, for example, are especially strongly emphasised in the regulatory culture of the UK and yet are relatively inflexible and narrow in their scope, closed to divergent values and framing assumptions and hubristic about uncertainty. In short, such approaches aspire to the dubious status of the 'analytical fix' tending to deny the essentially subjective and political aspects of the appraisal of technological risk.[22] Although the results of individual studies may be asserted with great precision and confidence, the picture across a range of such studies conducted by different bodies is typically far more variable and confused.[23]

Some of the participatory and deliberative procedures, on the other hand, are open to concerns over their verifiability and reproducibility and the transparency with which results can be linked with starting assumptions. There are questions over the extent to which such approaches may be considered truly 'representative' as well as fears that they may prove to be protracted or inconclusive in their findings. All approaches alike face difficulties of feasibility, efficiency and accountability as well as raising issues of manipulation and 'capture'. Here, a balance must be struck between overly prescriptive methods whose spurious precision can foster costly errors and unduly ambiguous approaches which fail to allow the drawing of any meaningful conclusions.

Emanating from unlikely origins in military logistics and operations research in the Second World War, a family of techniques which might collectively be known as multi-criteria appraisal have been developed over the years in the wider field of decision analysis.[24] In the past two decades, these have reached a stage of some maturity.[25] Informed by an often highly technical literature on rational choice and utility theory,[26] there has been a tendency for such techniques to become increasingly complex. Nevertheless, they are employed in many forms, to differing degrees and with varying success in certain countries, especially in fields such as transport and land-use planning,[27] siting,[28] energy policy,[29] waste management,[30] medicine,[31] commercial decision making [32] and sometimes technology assessment.[33] They are especially well-established in a public policy context in northern European countries such as Denmark and the Netherlands,[34] but – outside of the field of management science – seem considerably less frequently used in the UK.[35]

The details of the various techniques of multi-criteria appraisal are well reviewed elsewhere.[36] Put briefly, the basic procedure at the heart of the present methodology involves multiplying a performance ***score*** under an individual appraisal criterion with an

15 NRC, 1996
16 Eg: Lee, 1989; Wathern, 1991; OECD, 1993; van den Berg et al 1995.
17 Eg: Keeney et al, 1976; Janssen, 1983; Nijkamp et al, 1990; Bogetoft and Pruzan, 1991.
18 Eg: Covello et al, 1985; Suter, 1991; Royal Society, 1992.
19 Eg: articles in International Journal of Technology Management, 11(5/6), 1996; Rip et al, 1996.
20 Eg: Pearce and Nash, 1981; OECD, 1989; Pearce and Turner, 1990; Cropper and Oates, 1991.
21 Fiorino, 1989; Sclove, 1995; Bohmann, 1996; Webler, et al, 1995; Durant and Joss, 1995; Renn et al, 1996.
22 Stirling, 1998, 1999.
23 Stirling, 1997.
24 Keeney et al, 1976; Bell et al, 1977; Starr and Zeleny, 1977; Rivett, 1980; Fischoff et al, 1980; Edwards and Newman, 1982; Chankong and Haimes, 1983; Winterfeldt Saaty, 1988
25 Bana e Costa, 1990; Borcherding et al, 1990; Clemen, 1991; Janssen, 1994.
26 These issues are discussed in more detail by Kelly (1978), MacKay (1980), (Collingridge, 1982), Bonner (1986) and Bezembinder (1989) with a convenient summary of the discussion provided by Pearce and Nash (1981).
27 Friend and Jessop, 1977; Pinkus and Dixson, 1981; Voogd, 1983; Nijkamp et al, 1990; Bogetoft and Pruzan, 1991;
28 Kirkwood, 1982; Keeney and Nair, 1977; Merkhofer and Keeney, 1987; Solomon and Cameron, 1985; Keeney, 1980
29 Hope and Owens, 1986; ORNL, 1989; Jones et al, 1988; Hope et al, 1988; Keeney et al, 1987; Kreczko et al, 1987; Stirling, 1997; Lubbers, 1989; van der Pligt, 1989; Hobbs and Horn, 1998; Climaco et al, 1988; Andersson, 1988
30 Renn et al, 1996
31 de Bruyn, 1990
32 Hax and Wiig, 1990
33 Field et al 1994; Sharif and Sundararajan, 1983; Phillips, 1989; Cetron, 1973; Japp, 1993; Covello, et al, 1985; Vlek and Cvetkovitch, 1989a
34 Janssen, 1994; Voogd, 1983; Nijkamp et al, 1990; Bogetoft and Pruzan, 1991
35 DETR, 1998; ILGRA, 1997; Chapman, 1981
36 Bana e Costa, 1990; Borcherding et al, 1990; Clemen, 1991; Janssen, 1994; Salo, 1995

importance ***weighting*** assigned to that criterion. The overall ***rank*** derived for each option is therefore simply the weighted average of these scores.

There are, of course, a number of additional features of the methodology required in order to ensure consistency. These are discussed later in this report. For present purposes, however, the central point is that multi-criteria procedures are often employed in a similar fashion to orthodox cost-benefit or risk analysis as a way of determining a single 'objective', 'rational' or otherwise 'definitive' solution to an intractable social decision-making problem. There is nothing intrinsic about a multi-criteria methodology which prevents its use in an attempt to impose a spurious 'analytical fix' on a complex political controversy. A recent example of the unsuccessful use of such techniques in this regard in the UK is provided by the site selection procedure for a national radioactive waste repository undertaken by the nuclear industry in the late 1980's.[37]

However, in reviewing this and other similar applications to complex and controversial public policy problems, it is evident that when reduced to their most straightforward form the techniques of multi-criteria appraisal offer the *potential* for a radically more flexible decision aiding tool. A positive example of the use in the UK of a multi-criteria approach in the context of a more open, deliberative appraisal process is provided by recent work by the Environment and Society Research Unit at UCL for the Environment Agency on Local Environmental Action Plans.[38] Instead of being used in the narrow and often counter-productive fashion of an 'analytical fix', this work shows that multi-criteria techniques can be used instead as a 'heuristic' a way of exploring the key dimensions of a risk issue and establishing their characters, relationships and relative importance.

In short, rather than using them to rationalise a particular uniquely 'objective' or otherwise ostensibly definitive position, multi-criteria appraisal methods can instead be used as a way of 'mapping' a risk debate. Here, the explicit separation of the concepts of relatively technical 'scores' and more openly subjective 'weightings' (an idea common to all multi-criteria approaches) constitutes an especially important feature and establishes a significant precedent for the treatment of other dimensions of appraisal. Factors such as the scope of analysis, the framing of crucial assumptions and the treatment of uncertainties can also be handled in multi-criteria appraisal in a relatively open, transparent and systematic fashion. In this way, the use of multi-criteria appraisal as a decision 'heuristic' offers a way of establishing the main contours in a risk debate and of clarifying key areas of dissent and convergence between different constituencies. When used in this fashion, a multi-criteria technique may be distinguished from the more conventional 'analytical fix' by referring to it as a 'multi-criteria mapping' (MCM) approach.[39]

The particulars of the MCM method adopted in the present study are fully described in Chapter 3 and Annex 3 of this report. Since many of the more elaborate complications in other multi-criteria techniques are introduced in an attempt to justify the unitary prescriptive conclusions of the 'analytical fix' they are superfluous in a heuristic 'mapping' approach. Indeed, the justification for such complexity can even be questioned in its own right, since none of the many elaborate techniques developed in multi-criteria evaluation over the past four decades may claim fully and finally to have resolved the fundamental theoretical problems encountered in the social appraisal of risk.[40] The guiding principle chosen for the present exercise, then, has been to employ the simplest of all possible theoretically-valid approaches. This raises a series of detailed methodological points which are also discussed in further detail in Annex 3.

Before embarking on an account of the present pilot study, however, it might be useful clearly to summarise the basic properties which it is hoped that an MCM approach might be held to display when compared with other techniques for the appraisal of technological risk. Some of the more important qualities which may be seen as more or less desirable, achievable or practicable in different contexts are given in the left hand column of Table 1. These constitute the criteria under which the value of the present exercise might itself be appraised. The corresponding aspirations that might be entertained on behalf of an MCM approach are set out in the right hand column of Table 1.

Put simply, the hypothesis here is that in its most straightforward form an MCM approach may hold the *potential* to address (at least to some extent) *all* the concerns raised above in such a way as to foster the virtues both of participatory deliberation *and* of systematic analysis. If this is the case, then multi-criteria approaches in general, and MCM in particular, appear to be worthy of greater attention in the debate over risk assessment and technology appraisal than is currently the case in the UK. Either way, there is a need for exploratory work aimed at establishing the extent to which a straightforward MCM heuristic may yield practical benefits as a possible complement to other analytic and deliberative approaches in the conduct of appraisal for the regulation of new technologies such as genetically modified crops. It is with the objective of testing this hypothesis that the present pilot study has been undertaken.

37 NIREX, 1995; Stirling, 1996. This said, it must be acknowledged that this was as much a feature of the presentation as of the conduct of this particular analysis and that it was a strength of the multi-criteria methodology employed in this case that the critique could be so readily articulated.

38 Clark et al, 1998.

39 This term was introduced in the context of a hypothetical demonstration exercise in Stirling, 1997.

40 Watson, 1981, Stewart, 1996, Stewart, 1992, Vlek and Cvetkovitch, 1989b, Bezembinder, 1989; Smith, 1992; Vansnick, 1986; Gonzalez and Tversky, 1990; Bana e Costa and Vincke, 1990; Collingridge, 1982; Arrow and Raynaud, 1986; Salo, 1995

2.3 Relevance of GM as a subject matter

The controversy over the introduction of genetically modified (GM) crops and foods in Europe is currently a highly topical and controversial risk issue which is taxing the minds of industry, regulators and society in general. The advent of GM strategies in agriculture opens up a new arena for discourse over technology and environment. There is general agreement that there exists at least the potential for serious, irreversible harm. However, there is considerable scientific uncertainty over the form and magnitudes of the possible effects and, as yet (by

Table 1:

Some generally desirable properties in the appraisal of technological risk as a set of aspirations in a multi-criteria mapping approach.

QUALITY	DEMAND	MULTI-CRITERIA MAPPING
Flexibility and breadth of scope	No artificial constraints should be imposed on the type of issue that can be taken into account in appraisal simply because of the nature of the chosen method or metric for measuring risk.	MCM can be used with a range of quantitative or qualitative methods. This means that it is relatively unconstrained and involves no necessary emphasis of certain types of factor over others.
Openness to divergent choices, values and framing assumptions	It must be possible to include and articulate a variety of different interests, values, priorities and assumptions and a range of choices of alternative technological or policy options.	Divergent perspectives are expressed in MCM through open-ended approaches to 'choice options', 'appraisal criteria', 'performance scores' and 'importance weightings'.
Candour about uncertainties	There should be no undue constraints concerning the nature or scope of the uncertainties and analysis should explore a wide range of different possible outcomes.	MCM permits sensitivity analysis and probabilistic modelling (where appropriate), but also admits the unconstrained consideration of different possible outcomes.
Heuristic 'mapping' of performance	Appraisal techniques should be treated as knowledge-gathering 'heuristics' rather than as prescriptive 'analytical fixes' which determine in themselves a single , apparently definitive, 'rational decision'.	Sensitivity analysis of different 'framing assumptions' in MCM allows systematic examination of the links between different perspectives and the associated 'reasonable decisions'.
Analytical discipline and rigour	The techniques should be theoretically well-founded and systematic in their execution and should be repeatable and verifiable in practice.	MCM is founded in the well-established disciplines of rational choice and utility theory, with an extensive literature developing principles of good practice.
Transparency to review	The techniques should allow for an 'audit trail' explicitly linking the results with the various inputs, assumptions and parameters adopted in analysis	Participants and third parties can review the treatment in MCM of crucial determinants such as 'options', 'criteria', 'scores' and 'weightings', verify results and explore sensitivities.
Openness to participation	The techniques should be consistent with the existing aspirations and trends towards more open, participatory, deliberative approaches to regulatory appraisal.	MCM requires as inputs both technical information and intrinsically subjective framing assumptions. It therefore necessarily involves both expertise and citizen deliberation.
Feasibility and efficiency as part of a regulatory process	The techniques should not be too demanding, expensive to implement, unduly protracted, ambiguous in their implications or lacking in robustness.	Multi-criteria approaches are widely applied in fields such as energy policy, landscape planning, siting and priority-setting in health care. The modest scale of the present exercise is also an indication.

contrast with chemical or nuclear risks), little accumulated practical experience to draw upon. This has led to the evolution of a set of controls which are intended to be precautionary in nature where it is accepted that action to avoid harm may be taken in the absence of scientific proof with the conduct of risk assessment being required *before* experimental or commercial use of a particular genetically modified organism is allowed.

Despite this proactive, cautious approach to risk regulation for instance enshrined in the European Commission's Deliberate Release Directive (90/220/EC) the regulatory appraisal process has failed to gain confidence, either of non-governmental organisations (NGOs), private industry[41] or the general public.[42,43] This lack of confidence arises because, among other things: the scope of the regulatory appraisal is disputed; there is a general lack of trust in official reassurances of safety (particularly in the wake of BSE); and perceived benefits are not explicitly included in the evaluation process. Industry and regulators have expressed frustration in the belief that the precautionary approach is being used to demand an unrealistic absolute proof of safety.

It has also been almost impossible to gain agreement between European Member States over whether particular commercial releases of GM crops are environmentally 'safe', despite a supposedly common approach to their risk assessment.[44,45] Disputes routinely emerge over the appropriate scope of risk assessment. Even where there is agreement over the possibility that effects will occur, notions of what constitute *adverse* effects remain strongly contested.

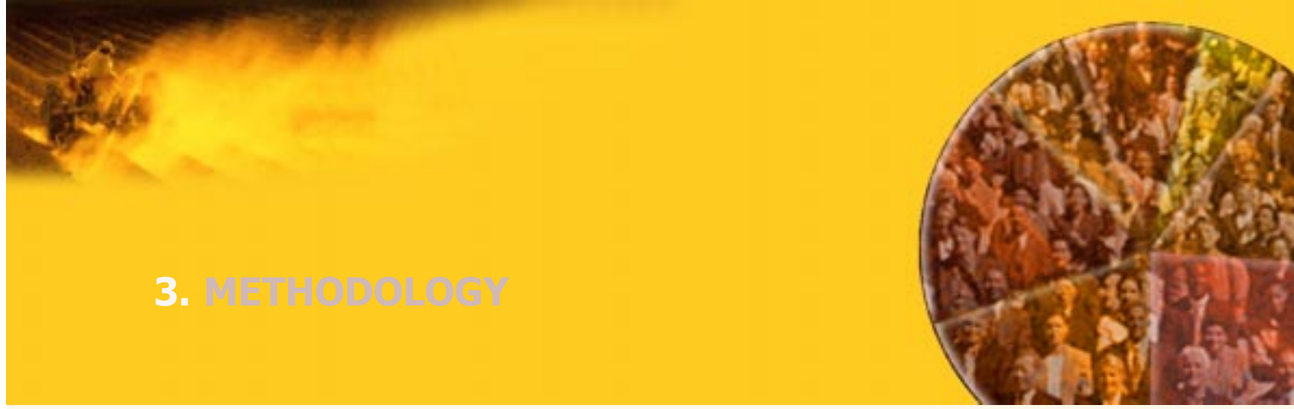
These sorts of problems with the risk assessment of GM crops, are typical of those which beset the use of conventional risk assessment and cost benefit analysis in other areas. Accordingly, it may be that MCM can offer a way of addressing some of the associated issues. The present MCM pilot study of a GM crop was chosen not only as a highly topical case study, but as a way of exploring in a practical setting the type of information which an MCM might provide and whether this might provide the kind of knowledge which would assist in the taking of more robust, socially informed decisions than are permitted by current risk assessment procedures. Because the European Commission has made proposals to revise the Deliberate Release Directive and the UK Government is reviewing the arrangements for making decisions about biotechnology, it is hoped that the findings of this pilot study might also help to inform these deliberations.

2.4 The Background to the Present Project

The present project evolved out of a series of roundtable meetings between industry and NGOs organised by Unilever and the Green Alliance. In particular, it takes forward one of the recommendations of an earlier report which was part of the same process and conducted by the Centre for the Study of Environmental Change at Lancaster University, *'Uncertain World. Genetically Modified Organisms, Food and Public Attitudes in Britain'*. This report urged that there should be a "*programme of institutional experiments, aimed at greater involvement of the public, in order (a) to develop more socially resilient shared understandings of the conditions of acceptability (or otherwise) of GMO foods, and (b) to improve the 'social intelligence' of industry and Government vis a vis relevant public understandings*". It also builds on the Unilever, Sainsbury's and Consumers' Association *'Confronting Risk. Finding new approaches to risk'* seminar in October 1997 which identified a "*need to establish a wider knowledge base to decisions and to institutionalise reflection and feedback so that decisions can be continually reviewed in the light of changing circumstances.*

The project was co-ordinated by GeneWatch and funded by Unilever. The research was undertaken between June 1998 and May 1999 by Sue Mayer of GeneWatch and Andy Stirling of SPRU, the centre for science and technology policy research at the University of Sussex.

41 Mayer et al, 1996
42 EPCAG, 1997
43 Grove-White et al 1997.
44 Von Schomberg, 1998.
45 Wynne and Mayer, 1999.



3. METHODOLOGY

3.1 Outline of the Approach

The flow chart in Figure 1 briefly summarises how the present pilot MCM study was implemented. The various steps are described in more detail in the following sections. Although the multi-criteria techniques have been widely used in the appraisal of technologies in other fields (especially outside the UK and particularly in the energy sector) this is the first application of this type known to the authors in the field of agricultural biotechnology. In applying such a relatively novel approach in a newly emerging arena, it was decided to restrict the study to a pilot scale in order to allow careful consideration of whether and under what circumstances the technique might be useful. The number of participants was therefore restricted to just twelve people, all of whom are intimately familiar with the issues involved.

3.2 Choosing the Subject

Genetically modified herbicide tolerant oilseed rape was chosen as the subject for this pilot MCM both in order to provide a concrete focus and because this is a 'real' topical development currently under intensive scrutiny. Although GM herbicide tolerant oilseed rape is the specific subject of the inquiry, it was placed in the setting of alternative options for the production of oilseed rape. So, as will become clear from the account of the conduct of the MCM, the intention was not to make a specific pronouncement on the safety, general desirability or otherwise of GM herbicide tolerant oilseed rape, but rather to evaluate its relative performance under different perspectives. By comparing the kind of information generated by such an approach with that from a conventional risk evaluation it becomes possible to investigate whether MCM can be useful in these practical situations. As a result, the outcome of this MCM cannot be simply or uncritically extrapolated to any other (still less, all) GM crops.

3.3 Selecting the Participants

The twelve individuals who agreed to participate in this study were approached on the basis of their established positions as representatives of leading protagonists in the current UK debate over the

Figure 1

Flow chart of MCM technique and how applied in this pilot study.

DECIDE SUBJECT AREA
- genetically modified, herbicide tolerant, oilseed rape

▼

DEFINE BASIC POLICY OPTIONS
- No GM crop, organic agricultural system
- No GM crop, integrated pest management system
- No GM crop, conventional agricultural system
- GM crops with segregation and labelling
- GM crops with post-release monitoring
- GM crops with voluntary controls on areas of cultivation
- Up to six to be chosen by the participant (these could include combinations of the above)

▼

SELECT PARTICIPANTS according to:
- sector of debate
- relevance of expertise
- spread of opinion

▼

INDIVIDUAL INTERVIEWS (2-3 hour session)
- select additional options
- define criteria by which to evaluate
- score options under each criterion, specifying uncertainty when relevant
- decide relative weighting of criteria

▼

ANALYSIS (Qualitative and Quantitative)
- group criteria
- identify areas of agreement/disagreement
- examine uncertainty patterns
- conduct sensitivity analysis
- investigate diversity

▼

FEEDBACK ON PRELIMINARY RESULTS
- participants reassess or confirm initial input
- results and analysis adjusted where necessary

▼

DELIBERATION
- discussions between participants on the basis of adjusted results

▼

FINAL ANALYSIS AND REPORT

development and regulation of GM technologies in the field of food production. As such, each participant holds (albeit from different perspectives) a strong general knowledge of the issues raised in contemplating GM strategies and their alternatives, as well as specialist expertise on certain aspects of these issues. Both as individuals and in their institutional context, the selected group of participants may be considered to be significant 'actors' in the policy arena and it would not be surprising to find any of them on a government advisory committee in some professional capacity.

As can be seen from Table 2, the organisations and constituencies with which the various participants are associated all maintain active interests in the subject of GM crops and all stand to be affected in different ways by their introduction. The group as a whole spanned a diverse range of institutional interests and perspectives (ranging from strongly favourable to strongly opposed to GM strategies). Indeed, it is rather unusual in the UK to find such a disparate array of contending interests co-operating in an individual appraisal exercise of this sort. Indeed, this ability to secure wider trust and involvement may, in itself, count as a particular feature of an MCM approach. However, despite the range of expertise and spread of opinions, the resulting perspectives should not be thought of as being representative in any formal sense of those extant in the wider society. It would take further research beyond the scope of the present pilot study to determine how representative (or otherwise) of wider public attitudes were the perspectives adopted by these particular participants.

Table 2

The participants

AREA	CODE
Academic scientists	C, J
Government safety advisors	E, F
Religious and public interest groups	A, D, G, I
Agriculture and food industry	B, L, H, K

Each participant was supplied with written information about the project and their agreement to be involved in the project obtained. Due to the high profile and controversial nature of the issue, the novelty of the present pilot exercise and the sensitivities of the different protagonists it proved possible to secure full involvement only by ensuring both individual and institutional anonymity. In order to respect this condition, each participant has been assigned a letter (see Table 2) and their particular institutional affiliations are not identified. These code letters are used in the analysis and in the presentation of results.

3.4 Defining the 'Basic Options'

One crucial methodological difference between this MCM and a conventional risk assessment is the way in which the central question under scrutiny was constructed. Rather than asking whether an individual course of action is 'safe', 'unsafe' or even 'safe enough' (as is common in regulatory risk assessment), this MCM takes as a starting point a series of different possible choices and seeks to determine the relative performance of these 'options' in relation to each other under a range of different criteria.

In this pilot study, the principal focus concerned the relative merits of GM herbicide tolerant oilseed rape compared with different technical and policy options for the cultivation of oilseed rape and the performance of all these options relative to one another. In order to ensure some degree of comparability between the perspectives taken by the different participants, six 'basic' policy options were identified and defined in advance by the researchers. All participants were asked to consider and appraise these six options. In order to ensure that the analysis was not unduly constrained or biased by this externally imposed framework, participants were able to add up to six further options which they were entirely free to define as they thought most appropriate (see Table 3).

The geographical context for the six basic options is that of the UK. Some of the chosen options are - to varying extents – somewhat hypothetical. For example, no organic oilseed rape is presently under production in the UK because the current processing system uses a chemical (hexane) which is not permitted in organic food production. Likewise, all the options under discussion are to some degree or another - somewhat stylised. Categories such as 'organic', 'integrated pest management' (IPM) and 'conventional' apply across a variety of practices and contexts. However, such is necessarily the case in any practical comparative appraisal. Between them, the basic options under consideration in this exercise encompass a wide range of possible strategies and so provide a potentially useful frame of reference in current debate over the introduction of GM foods.

46 Proposal for a European Parliament and Council Directive amending Directive 90/220/EEC on the deliberate release into the environment of genetically modified organisms (COM (98)0085 C4-0129/98 - 98/0072 (COD)).

47 SCIMAC 1998.

The basic GM options were chosen to reflect the three main approaches to regulatory control which are currently in place or being contemplated. Labelling on the basis of the presence of foreign DNA or protein is the approach taken in the EC's Novel Foods Regulation (258/97). A mechanism for post-release monitoring is one of the revisions to the Deliberate Release Directive being proposed by the European Commission.[46] Voluntary controls on areas of cultivation are being contemplated in the development of industry guidelines on the growing of herbicide tolerant crops.[47]

The three basic non-GM options reflect the active debate over contending agricultural strategies and environmental protection especially in relation to herbicide tolerant crops, how they will affect the pattern of chemical usage and what effect this may have.

As a result of the choice of these six 'basic options', the approach adopted in this exercise allowed consideration of a wide variety of agricultural strategies (both including and excluding use of GM technology) together with a range of different regulatory mechanisms for the control of genetically modified crops. The issues raised in the comparison of the performance of these 'basic options' transcend the bounds of narrow 'safety' concerns alone. As a result, the kind of information that may be expected to emerge from a question framed in this way is very different from that generated by the orthodox risk assessment of individual options on a case by case basis.

For example, by directing attention at the alternatives, light may be cast on the consequences (whether positive or negative) of *not* pursuing the GM options. Likewise, the scrutiny of a range of alternative possible regulatory strategies might reveal the type of conditions under which GM crops would be acceptable under different perspectives. The inclusion among these 'basic options' of possible strategies which specifically rule out the use of GM crops was an important dimension to this project, since it ensured that the appraisal was not framed from the outset in such a way as to prejudge the nature and implications of the results. Furthermore, by allowing participants to specify additional options, the possibility was raised that alternatives might emerge which would merit further consideration by others. It is likely that this broad and systematic approach to the framing of the exercise which allowed the securing of participation from such an unusually diverse group of organisational interests.

For the purposes of clarity in evaluation, it was assumed that each individual 'basic option' was pursued to the exclusion of all others in the UK. In this way, the assessment of individual options was not complicated by questions over potential interactions with other options. The question of the pursuit of *diverse mixtures* of options in the UK was introduced as an additional factor in appraisal in a later stage of the project.

3.5 The Interview Process

The twelve participants were interviewed on an individual basis between June and September 1998. Interviews lasted between 2 and 3 hours and were tape recorded in order to provide a detailed record

Table 3

The definition of the 'basic options' appraised by all participants

OPTION	DEFINITION
Organic Agriculture	All farming and food production conducted under present day organic standards
Integrated Pest Management	All farming and food production conducted using systems designed to limit but not exclude chemical inputs and with greater emphasis on biological control systems than conventional systems.
Conventional Agriculture	All farming and food production conducted under present day intensive systems.
GM oilseed rape with segregation and present systems of labelling	Labelling based on the presence of foreign DNA or protein in the final product.
GM oilseed rape with post-release monitoring	Monitoring for effects (mainly environmental) conducted on an on-going basis after commercialisation.
GM oilseed rape with voluntary controls on areas of cultivation	Areas of growing of GM oilseed rape restricted on a voluntary basis to avoid unwanted effects such as gene-flow and cross fertilisation of non-GM crops.
Up to six additional options to be specified by participant	Any option of participant's choice including combinations of the above if desired

of important verbal descriptions and qualifications for later consultation. One researcher (SM) attended all the interviews to ensure continuity and comparability of interpretation. The other researcher (AS) attended five. During the interview, a four stage iterative process was undertaken, comprising: (i) the identification of additional options; (ii) the specifying of appraisal criteria under which the options should be assessed; (iii) the scoring of the performance of each option under each criterion and (iv) the weighting of each criterion in terms of its relative importance. Figure 2 provides a schematic representation of this process. The essential iterative and reflexive properties of the process meant that participants were able to return and include further options or criteria during the interview if, as things developed, they thought of others they would like to add.

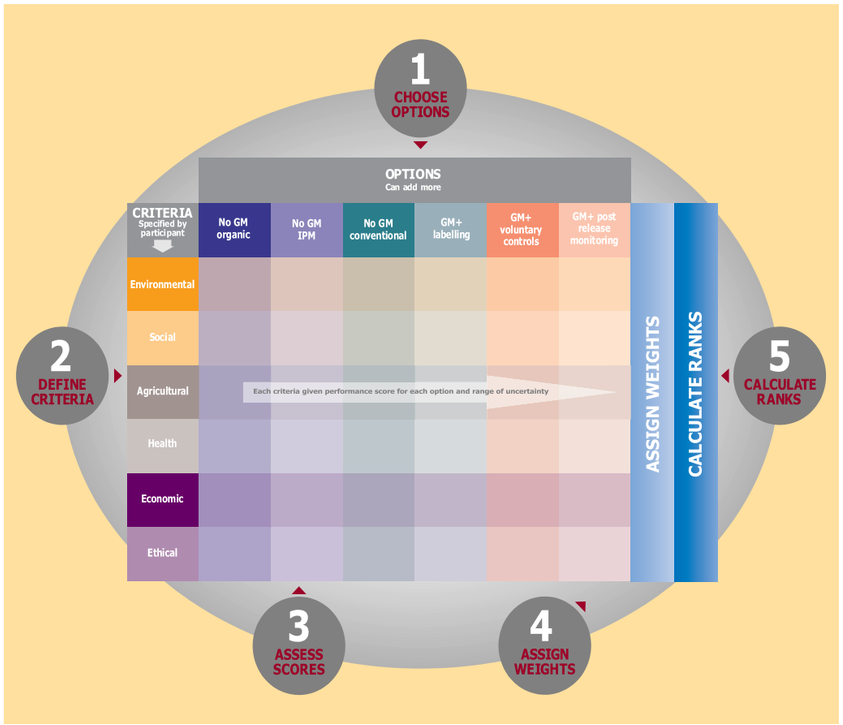
Options

The first step in the interview process was to discuss the specific definitions for the six basic options under appraisal. Participants sought to clarify aspects of these definitions which might have a bearing on their own appraisals. Depending on judgements over the completeness or resolution provided by this set of basic options, participants then defined up to six additional options of their own choosing.

Criteria

Participants were then asked to define a maximum of twelve criteria which they would use to evaluate GM herbicide tolerant oilseed rape and the production of oilseed rape in general. There was no restriction placed on the scope or form of criterion which a participant could specify. However, although different criteria might be related in various ways, each must be considered 'independent' in the sense that the associated assessments of performance do not *depend* on judgements of performance under other criteria. Participants were asked to describe in as much detail as possible what each individual criterion meant to them, specifying precisely, for instance, what they meant by any broad, general terms such as 'sustainability', 'precaution' or 'efficiency'. Because there might always be more than twelve criteria under which oilseed rape might be evaluated, participants were asked to concentrate on those they thought were *most important* in the evaluation.

Figure 2 The Multi-Criteria Evaluation Process



Again, if new issues emerged as the appraisal procedure progressed, participants were free at any stage to resolve new criteria and assess their options under these as well.

Scoring

Having specified their appraisal criteria, the participants were asked to score each policy option under each criterion. This is the part of the MCM exercise which deals with the 'technical' side of appraisal. Participants were thus asked to justify the scores which they assigned under their various criteria by reference to what might be called 'scientific' or 'technical' considerations, drawing on their own knowledge and expertise as established professionals in the broad field of food policy and as representatives of important institutional protagonists in the wider social debate over the use of GM crops in food production.

In principle, participants could express scores in terms of the established measuring units appropriate to each individual criterion (such as tonnes of herbicide used, numbers of species affected or monetary values). Where such metrics are not felt to be available or appropriate under particular criteria, however, relatively high, medium, or low performance can be expressed in MCM simply by adopting an arbitrary cardinal scoring scale (such as 1 to 10,or 1 to 100). In such cases, explicit 'anchor points' are established for the assigning of scores, for instance by reference to the current *status quo* as a mid-range score or *zero risk* as a maximum score. In all cases a high numerical score corresponds with high performance and *vice versa*.

In practice, scoring was performed by proceeding down the list of options under each individual criterion. The performance score assigned to each option was arrived at by what was often an intense and demanding process of systematic and iterative deliberation, making reference to the performance of all other options under that criterion. Here, reference would typically be made in discussion to a wide variety of conditioning assumptions and countervailing factors. The scoring exercise also provided a means to check that the different criteria as defined by each participant were mutually independent for practical purposes.

Participants were asked to assign both high (optimistic) and low (pessimistic) scores for each option under each criterion. This procedure allowed participants to express their judgements over the importance of technical uncertainties and case-by-case or context-dependent variability, where appropriate. Where neither uncertainty nor variability were felt to be a factor, the pessimistic and optimistic scores could be identical. Participants were also asked to describe the 'framing assumptions' which they were applying in each case such as their confidence in good practice or regulatory regimes or the assumptions they were making concerning dynamic changes over time.

A lap-top computer running a simple procedure written by one of the authors (AS) for Microsoft Excel 97™ proprietary spreadsheet software was used in order to perform a straightforward 'linear additive weighting' multi-criteria procedure. Essentially, this involves simply taking the performance scores assigned by the participants and multiplying them by importance weightings which are assigned separately in the next stage of the exercise to express the relative priority attached to the different criteria. The result is a 'ranking', reflecting the overall performance of each option under all the criteria taken together, taking account of the relative importance of these criteria under the perspective in question.

The spreadsheet automatically 'normalises' the scoring scales to preserve the ratios while avoiding inadvertent bias due to the arbitrarily higher numerical values which might be employed under some scoring scales compared with others. For instance, without this kind of correction, a score expressed in tonnes per year would otherwise apparently differ by a factor of one thousand from the same score expressed in kilograms per year. Likewise, scoring rated on a scale of 1 to 100 was rendered comparable with scoring rated on a scale of 1 to 10. The results of the scoring process were displayed by the computer in real time for each participant during the session as a bar chart. The bar charts showed the rankings of options under both 'pessimistic' and 'optimistic' performance scores.

Weighting

The final step in the interview process was the assigning by each participant of numerical weightings to reflect the relative importance of each of their appraisal criteria. By contrast with the 'technical' and 'scientific' considerations addressed in scoring, this is the stage in the exercise when explicitly subjective value judgements are made. The weightings reflect how much participants *care* about the differences in option performance under each criterion.

It was explained to participants that the weightings which they assigned to their criteria should not be thought of in an isolated abstract sense. For each participant, the assigning of weightings was inextricably linked to the particular scores which they had allocated to the various options under each criterion. For instance, the relative importance of

'cancer risk' cannot be compared with that of 'rural employment' unless it is clear 'how much risk' and 'how much employment' are involved. For this reason, the computer model automatically identified and represented the scores determined for the best and worst option under each criterion, allowing the participants more readily to consider the practical implications of their weighting judgements.

Where the difference between best and worst option under one criterion is judged to be twice as important as the difference between best and worst option under a second criterion, then the weighting assigned to the first criterion will be twice that assigned to the second. The final weighting scheme is a set of numbers whose ratios reflect the relative importance of scoring differences under the various criteria.

By contrast with many multi-criteria exercises, participants in the present study were left relatively free to undertake the weighting process in whatever way they felt most comfortable, with the interviewer providing guidance and suggestions where requested. Starting from a default position where equal weighting was assigned to each criterion, participants usually began by ordering the criteria simply in sequence of their relative importance. The *intensity* of the differences in importance were then addressed by altering the ratios of the weightings one by one. This continued in an iterative fashion until a final set of numbers was arrived at which the participant felt comfortable with.

The arbitrary weighting numbers entered by participants were recalculated and represented by the programme on the lap-top computer in simple percentage terms. This corresponded with an intuitive model of importance weighting in terms of the sharing of 100 'importance points' across the various criteria. The computer also displayed as a bar chart the consequences for the overall ranking of options of each change made to the weighting scheme. All participants had access to the computer at the end of the process and were able to manipulate the weightings themselves in order to explore sensitivities. The weighting procedure was not concluded until each individual participant expressed satisfaction that they had arrived at a meaningful expression of their position. Although not all made use of the computer in this way, all agreed that they were happy with the outcome.

Since they all had access to computers and the interface with the computer model was rather straightforward and intuitive to those who had undertaken the exercise, all participants were offered the option of retaining their own results on disk as a customised Excel™ spreadsheet in order to allow them to explore the consequences of different weightings at their leisure. Four participants took up this offer but in no case did this result in any changes to the weighting scheme arrived at during the interview itself.

3.6 Preliminary Analysis

Following the round of interviews, a preliminary analysis of the results was conducted. This included: (i) the grouping of criteria, (ii) a systematic sensitivity analysis in order to examine the effect of increasing and decreasing each participant's criteria weighting values and (iii) an exploration of the effect of introducing a degree of diversity into the mix of options, based on the rankings arrived at by each participant.

Grouping Criteria

The total set of appraisal criteria reflect a wide range of considerations viewed from a disparate array of perspectives. The interviews revealed that even where individual criteria adopted by different participants are apparently similar the way these criteria are framed in the process of determining scores may differ quite radically between participants.

However, there remained some scope for the grouping of criteria into a number of broad general categories for the purposes of exploring overall patterns. With this aim, the 117 individual criteria developed by participants (see Section 4.3) were ordered by the researchers into six groupings: 'Environment', 'Agriculture', 'Health', 'Social', 'Economic' and 'Other' issues. These categories were not established in advance of the interview process, but were developed relatively 'inductively' during the preliminary analysis on the basis of the criteria actually selected by participants. They are listed in more detail in Table 5 in Section 4.3.

These six groupings of criteria are rather 'conservative' in nature, tending to reflect the categories of issue which are most commonly recognised in the wider policy discourse. Other approaches, for instance, might have been to categorise criteria according to their 'scientific', 'technical', 'ethical', or 'political' (to do with agency or control) content. However, this would have been much more difficult to do, given the way in which the participants actually described their own criteria. Nevertheless, it remains the case that many such cross-cutting issues are intertwined in the six groupings displayed in Table 5.

Within these six broad groupings, a series of more specific sub-groupings of criteria were also resolved (Table 5) as an illustration of the kind of issue raised in the appraisal. Due to the many instances of overlaps and inconsistencies of framing between participants, these sub-groupings of criteria were not used in the formal analysis of results.

The group into which a criterion was assigned was determined by the definition articulated by the participant during the interview. In some cases, there was a degree of overlap across the six broad groupings. For example, one criterion was formulated as 'toxicity to wildlife and humans'. In such cases - which constituted only a minority of the criteria (7 out of 117) - that aspect which was emphasised during the interview was taken as the basis for categorising the criterion.

Following the interview stage and preliminary analysis, participants were asked whether they agreed with the way in which their criteria had been grouped. Except for minor amendments (which were adopted), no disagreements were expressed.

Sensitivity Analysis

Participants had already had the opportunity during interviews to experiment with changes to the weighting values which they assigned to their different criteria. However, the expression of the relative importance of different issues in simple numerical terms is a rather idiosyncratic (and to some an unfamiliar or even counterintuitive) process. Furthermore, the participant's initial weighting schemes were elicited at the end of a long and quite laborious interview process. Although all participants expressed satisfaction with their final weighting schemes during the interviews, the possibility could not be discounted that the assignment of weightings may in some cases have been unduly truncated by fatigue or pressure of time.

For these reasons, the preliminary analysis of results also concentrated on the systematic examination of 'sensitivities'. This involved an exploration of what the final rankings *would* have looked like for each participant if their weightings on each of the six groupings of criteria (ie: 'Environment', 'Agriculture', 'Health', 'Social', 'Economic' and 'Other') had been different by a factor of three either up or down. In other words, the weighting sensitivities were examined for each of the six groupings rather than for the individual criteria themselves (this would have been prohibitively complex both to perform and for the participants to interpret). Bar charts were generated which displayed the overall rankings (averaged over 'pessimistic' and 'optimistic' scores) obtained by the multi-criteria process for the different options under the original weightings and a threefold reduction and a threefold increase on this base. The overall difference between the lowest weighting and the highest weighting for each criterion explored for each participant was therefore a factor of nine representing a fairly considerable difference of possible views concerning the relative importance of the six broad groupings of criteria.

The sensitivity analysis was represented for each individual participant as a series of six bar charts (one for each criteria grouping) showing how the rankings obtained for each of their options might vary with changing judgements concerning the relative priority assigned to their criteria. No changes of weighting were suggested by any participant as a result of this further iteration in the process.

Diversity Analysis

The six basic options had been defined to preclude the parallel pursuit of other options in the UK. Given the complexity of the real world, it is obviously a rather artificial assumption that any individual option for the production of oilseed rape would be pursued in isolation. It is far more likely that a modest variety of different options would be pursued in different contexts. Such has long been the case, for instance, with regard to organic, integrated pest management and conventional intensive agriculture.

However, the property of diversity may be more important in appraisal than simply as a way of reflecting some of the practicalities of the real world. For instance, it is a matter both of common sense and a subject of recent theoretical inquiry that, when we face uncertainties which are as intractable as those widely acknowledged to attend the appraisal of GM crops and other long term agricultural strategies, one possible strategic response is to avoid "putting all the eggs in one basket". The deliberate pursuit of a small number of different options in parallel offers one potential way of attempting to reconcile the different values and interests relating to controversial issues such as food production in a plural democratic society such as that of the UK. In this sense, then, the deliberate pursuit of a diverse mix chosen from among the better-performing options identified under the perspective of each participant in appraisal might be seen as representing one element in a 'precautionary strategy'.

Based on work conducted in relation to the energy sector and in the field of evolutionary economics, a straightforward numerical index of diversity was employed in the preliminary analysis in order to allow for diversity to be considered as an additional 'criterion' in the multi-criteria appraisal. The index of diversity used (the Shannon-Wiener function) is a simple and relatively robust concept which is widely employed in disciplines such ecology and has recently been adopted by the UK government as a

general measure of diversity in the energy sector. The index was used to explore what *would* happen if progressively greater weighting were placed on diversity under the perspective of each participant. In other words, what sort of mix of options would result if each participant expressed an interest in jointly pursuing options *other than* that which ranked most highly under their own appraisal? The methodology and background are discussed in more detail in Annex 4.

A set of four pie charts were generated for each participant which showed what sort of mix of options would result if 'zero', 'low', 'medium' or 'high' weightings were placed on diversity among their chosen options and traded off against their other criteria. After briefly explaining the concept of diversity and the reasons why it might be interesting, participants were asked on a standardised feedback form:

Please consider the pie charts below, which show what a hypothetical mix of agricultural options for the production of oilseed rape in the UK might look like, IF the different options were relied upon to a degree related to their performance under your own appraisal. Each 'portfolio' represents a different judgement about the trade-off between, on the one hand, the benefits of diversity (as an acknowledgement of political pluralism, serious uncertainty or contextual variability) and, on the other, a wish to pursue only what is found (under your own appraisal) to be the best-performing option.

Please think about this issue of diversity and indicate on the attached sheet which, if any, of the hypothetical mixes shown below represents, in principle, the more acceptable scenario for you. We would also be interested in any general comments you may have on this question of diversity.

All seven participants who responded to the feedback form, responded also to this question.

3.7 Review by Participants

Each individual participant was sent a copy of their own results together with an anonymised list of all the options, criteria and weightings specified by others and the resulting option rankings. Based on the outcome of the provisional analysis described above, each participant was asked: (i) whether they were content with the way in which their criteria had been grouped, (ii) whether in the light of the results of the sensitivity analysis their weightings still reflected their opinions, and (iii) what opinion they had on the general issue of diversification across their various options ad which of a series of 'zero', 'low', 'medium' and 'high diversity' mixes (based on their own final rankings) they most favoured. Each participant was also asked to comment on the difficulty and utility of the MCM process as a whole and its individual parts.

Full responses to this review were received from seven of the participants. The other five felt unable to reply due to pressures from other work. A final meeting was held to review the results obtained in the study and discuss their implications, but similar scheduling constraints and pressures of work meant that only half the participants could attend. Here, there was a generally strong positive reaction to the project as a whole, and again no substantive changes of output or interpretation arose from this meeting

4. RESULTS

4.1 Engagement

For various reasons including lack of time, lack of information and, perhaps, lack of empathy with the approach (cf: see Section 5.1) two participants did not feel entirely comfortable fully engaging in one aspect or other of the multi-criteria procedure.

Participant G (from the religious and public interest area) felt uncomfortable assigning quantitative values either to the relative technical performance of options (scores) or to the relative importance of different issues (weightings).

Participant H (from the agriculture and food industry area) felt in need of more technical information before assigning performance scores, but did feel confident in ordering criteria in sequence of their relative importance (ie: as ordinal weightings).

In addition, Participant L (from the agriculture and food industry area) while comfortable assigning both scores and weightings made a distinction between criteria under which performance trade-offs might take place and criteria which would serve rather as 'hurdles' under which options would either 'pass' or 'fail', with failure leading to their complete exclusion.

This leaves a total of ten participants (more than eighty per cent) who completed all aspects the multi-criteria procedure. Despite this non-engagement by two participants in part of the quantitative procedure, the MCM process continued to yield useful information concerning qualitative considerations such as option definition and criteria choice for *all twelve* participants.

4.2 Options

The main focus of discussion in the initial stage of the interviews concerned the definition of exactly what was meant and implied by the short labels given to each of the different options. There was little discussion over the justification of the choice or specification of 'basic options', although there was some confusion about why non-GM options were included in an assessment of a GM crop. Several participants commented in the feedback and during the interview that they found the concept of comparing several options difficult to grasp, having been much more familiar with the assessment of GM crops on a case-by-case basis.

This finding may reflect the rather different approach embodied in current systems of risk assessment which look at the GM crop in isolation without comparing different policy options. Stepping outside such a framing into a different approach will inevitably be difficult and require time and experience. However, it may equally have been that the concept was not described adequately either during or before the sessions and this aspect would require more attention if the work was to be extended.

Nine of the twelve participants added a total of seventeen options to the list of core options (see Table 4). Adding or combining ***controls*** and/or making them ***compulsory*** were the most common type of additional options (7 of 17). Using GM crops inside integrated pest management or organic systems were the next most common type of additional option (5 of 17). Other were quality, public control, assessment of indirect effects and need.

Most of the additional options were described by participants as being candidates for their preferred or ideal option, although occasionally options were included simply because participants wanted to see how they would perform.

In four cases (participants B, C, J and K) one or all of the additional options performed as well as or significantly better than the basic options (Figure 3). These involved using a GM crop in an organic or IPM system, changing the decision making process and the *quality* of the final product. With the others, which concerned modifications to the GM options (participants A, F and I), there tended to be no (or only a marginal) improvement in the performance of the option with little impact on the overall ranking pattern. Participants G and H did not complete all stages of the MCM (see above) so their options could not be ranked.

The results suggest that, although controls on the use of GM crops are considered to be important,

how GM crops fit into the overall agricultural and food production system is more so. Therefore, one area of future research into the acceptability of GM crops may be to determine how, whether and under what circumstances GM crops could be part of organic or integrated pest management systems. However, because present organic systems explicitly rule out the use of GM crops, this is probably unrealistic.

4.3 Criteria

The twelve participants identified a total of 117 individual appraisal criteria, addressing a wide range of issues which they thought relevant to the assessment of the means of production of oilseed rape in general and GM herbicide tolerant oilseed rape in particular. Many of these criteria were on the face of it effectively identical to each other, others displayed differences of framing or emphasis of varying degrees of subtlety. Table 5 gives a summary of the criteria and how they were grouped by the researchers (and later approved by participants). The issues raised in the process of grouping criteria are discussed in Section 3.6 above. All the criteria are listed in Annex 1.

The definitions of many of the environment, agriculture and economic criteria included elements which could not be reduced to strictly technical or scientific parameters. For example, environment included aesthetic, ethical and visual criteria.

Table 4

Basic options and those added by participants

CLASS	OPTION	Participant and Option
Basic Options	No GM crop, organic agricultural system	All - 1
	No GM crop, integrated pest management system	All - 2
	No GM crop, conventional agricultural system	All - 3
	GM crops, with segregation and current system of labelling	All - 4
	GM crops with post-release monitoring	All - 5
	GM crops with voluntary controls on areas of cultivation	All - 6
Labelling and/or other controls	GM crops with segregation, current labelling and post-release monitoring	F - 7
	GM crops with segregation, full labelling and post-release monitoring and legally binding growing contracts	A - 7
	GM crops within controlled sectors (compulsory control)	A - 9
	GM crops with legally binding threshold for gene transfer to non-GM stream	A - 11
	GM crops with segregation and labelling according to means of production and source of gene, plus post-release monitoring	G - 8
	GM crops with segregation, comprehensive labelling based on process and generic restrictions on some classes e.g. in centre of origin	I - 7
	GM crops with segregation, full labelling and post release monitoring	H - 8
Agricultural system	GM crops, IPM system	J - 8
	GM crops, organic agricultural system, plus segregation, labelling and other regulations as required	J - 7
	GM crops, IPM system	G - 7
	No GM crops conventional and organic as now	K - 7
	GM crops in conventional and organic systems	K - 8
Assessment criteria	GM crops with assessment of indirect agricultural impact and assessment of need	I - 8
	GM crops with quality	B - 7
Other	Complete public control over choice	C - 7
	GM crops only in USA	A - 10
	No GM commodity crops	A - 8

Agriculture included farmers' rights, food stability and quality of life for agricultural workers. Economics included global economic considerations and sustainability. Even health, which seems to include the largest proportion of direct production-related criteria, also included nutritional impact and traceability. Neither were all the criteria tightly linked to the issue of genetic modification. Many were associated with the social and political ramifications of the adoption of the technology under certain conditions. These findings suggest that a broad range of non-technical considerations are felt to be relevant to assessments of a technological development even under headings which are conventionally considered to lie within the domain of technical expertise.

Agriculture was the most mixed group of criteria, including very specific agricultural practice issues, such as 'impact of the herbicide on managing tolerant 'volunteers' and some criteria which would have been categorised as 'social', had they not been specific to farmers (such as 'quality of farmers and agricultural workers lives'). This probably explains in part the different behaviour of this group of criteria in the later analysis. It also probably reflects the particular positioning of agriculture in the assessment. Whilst it has its own specific technical issues that demand consideration, other social issues are inextricably intertwined. The separation of agriculture issues from more general social issues may reflect participants' identification of the special effects which GM crops may have on all dimensions of agriculture and food production.

The social criteria were dominated by those to do with choice, control and agency. Seventeen of the twenty two criteria in this category could be thought of in this way. Interestingly these issues were raised only by those eight participants who were not part of either the production or formal evaluation process of the GM crop. This may reflect a sense of exclusion or perceived inability to influence the way in which choices are made. Either way, this finding merits further investigation.

The selection of criteria was evidently influenced by the professional interests and perspectives of the

Table 5

Criteria Groupings

Environment:	12/12 participants had at least one criterion addressing issues of:	
	sub-groupings:	biodiversity chemical use genetic pollution secondary or broader effects unexpected effects ethical, aesthetic and visual
Agriculture:	10/12 participants had at least one criterion addressing issues of:	
	sub-groupings:	weed control food supply stability agricultural practice
Health:	11/12 participants had at least one criterion addressing issues of:	
	sub-groupings:	allergenicity toxicity nutrition unexpected effects manageability
Economic:	10/12 participants had at least one criterion addressing issues of:	
	sub-groupings:	consumer price benefit farmers' or commercial users' benefit society benefit
Social:	8/12 participants had at least one criterion addressing issues of	
	sub-groupings:	individual choice, need, benefit and participation institutional demands social need, benefit and trajectory
Others:	4/12 participants had at least one criterion addressing issues of	
	sub-groupings:	ethics knowledge base

participants. For example, eight of the thirteen criteria selected by Participant A (from a public interest group) were concerned with health, consumer cost, choice, influence and information provision. In contrast, five of the six criteria selected by Participant B (from the agriculture and food sector) were concerned with farmers or commercial users' benefit. The other two individuals from the food supply chain (Participants L and H) had criteria similar to each other, covering the breadth of the different groupings and including broader concepts such as sustainability and the requirement for traceability or controllability, emphasising consumer confidence as part of their rationale for such criteria - issues which were relevant to their businesses. The Government safety advisers on environment and health (Participants E and F respectively) did not include any social criteria.

However, although participants did frame the issue by emphasising issues relevant to themselves, they also acknowledged other areas which had to be addressed. For example eight participants included criteria sub-categorised under environment as 'biodiversity' and eight also included criteria sub-categorised under health as 'toxicity'. So in addition to these generally agreed criteria, participants included more specific criteria according to their own interests.

Criteria also ranged from being very specific to very broad:

Could GM make the crop an invasive weed in the absence of herbicide" (Participant E)

Social welfare including cost, jobs, quality of life & occupational health" (Participant I)

Participants recognised that many criteria were aggregates of issues that needed 'unpicking' and were very complex, (such as 'the effect on the global economy' and 'biodiversity'), but felt they should legitimately be considered. Often criteria, although independent for the practical purposes of scoring, displayed close relationships, for example, 'transparency' and 'confidence in institutions'. These aggregated criteria, whilst seemingly impossibly complex, give important indications about the range of issues that individuals feel should be considered and areas that might need further investigation. Several participants also commented that their criteria might change if the MCM was repeated even in the short term as new issues emerged and others faded.

Three general findings emerge from this analysis of criteria definition. First, many criteria lie outside the scope of official risk assessments and for no participant is their whole range of criteria explicitly considered in the formal evaluation process of GM crops in the UK. Second, the choice of criteria is important in 'framing' the assessment and this reflects the values and interests of the people involved. Third, that both the direct effects and complex indirect consequences of a development are considered relevant in all domains of assessment.

4.4 Scoring

The scoring of options under the various criteria took up the longest period of the interviews. One person felt unable to score on principle, considering that all appraisal criteria are intrinsically unquantifiable in any meaningful way. One other participant was unable to score the criteria because they felt that they did not have sufficient knowledge to do so. Several participants pointed out that although *this particular example* (herbicide resistant oilseed rape) might score well or badly, their scoring for other GM crops might be very different.

In their feedback, four participants (E, K, I & J) expressed some kind of initial confusion with the process at this stage as did others during the interviews. This was generally associated with the inclusion of a wide range of options in the evaluation of a specific GM crop. Scoring under a criterion which was very specific to the GM herbicide tolerant crop for options which excluded GM was found by some to be conceptually difficult.

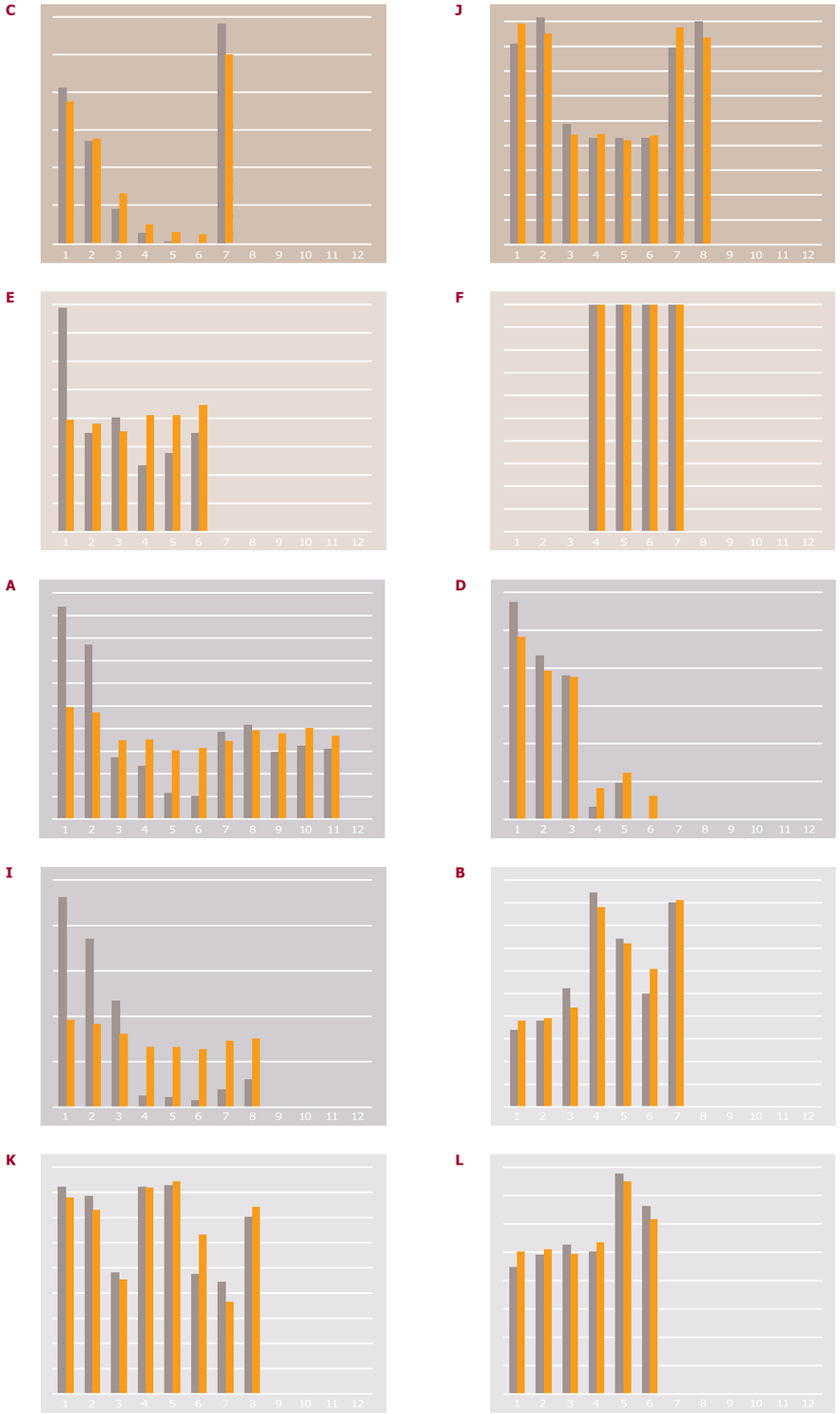
Those ten participants who did feel able to derive performance scores under their appraisal criteria used a numerical range such as 1-10 or 1-100. The scores were related to a particular baseline often the present day *status quo* so that scores could be determined according to whether performance under that criterion might be expected to get better or worsen. Given the nature of the present pilot exercise, no participant felt it appropriate directly to employ in scoring established metrics such as tonnes of pesticide, numbers of species affected or monetary values.

The importance of context

As participants justified their scoring during the interview session, they frequently either asked for definitions of the options or spelt out their own. It was clear that context was important and that the scoring of individual options was often influenced by how well the participant thought systems would work ***in practice.***

If you speak to enough top quality organic people you can see that as a system that can be quite good, if you don't you won't and likewise IPM can be terribly misused (Participant E)

Figure 3 The final ranks for ALL options for each participant



The average overall rankings for the basic options obtained under the perspective of each partici-pant (identified by their code letter). The orange and brown bars display the differences between the ranking orders over the various options obtained (respectively) under pessimistic and optimistic scoring assump-tions. They thus provide an idea of the impact of technical uncertain-ties on ranking orders. The first six pairs of bars represent the six basic options followed by the additional options added by each participant (see Table 4). Since the numerical values for the ranks are on an arbitrary linear scale, the axes for the pessimistic and opti-mistic scores have been scaled separately to allow easy com-parison of ranking orders.

So, for example, while there was general agreement that organic farming was beneficial for biodiversity (compared to conventional systems), just how beneficial this was, was seen to depend on how '*good*' the farmer was considered to be. A similar pattern was evident for the IPM option. Not only did opinions over likely practice influence the numerical values of the scores, they also influenced the uncertainties with which these scores were expressed. Expectations concerning the possibility for a range of good or bad practices or different environmental conditions dominated much of the discussion during the deliberation over scoring.

[weed control advantages] will vary from farm to farm because some people's land is more inherent to problems than others ..
(Participant B)

The importance of context was also seen when scoring the different GM options. There was some scepticism about the extent to which regulatory controls could manage risks:

Nine Perhaps, sorry, number six might not be if it's voluntary controls. Can you regulate for voluntary controls? You can regulate that there should be voluntary controls in place but you can't enforce them so that's probably a seven.
(Participant L)

Equally with the voluntary control, voluntary most people would be good, but there's always the rogue. (Participant D)

However, it was recognised that the exercise of controls may have effects beyond those intended. There could be negative or positive consequences . For example, with voluntary controls on area of cultivation, the reduction in area grown might restrict economic benefits.

And if you control the area of cultivation I think you probably add to the cost. (Participant F)

Labelling might have benefits in terms of recording and traceability as well as providing consumer information, leading to improved performance of this option over several different criteria.

[concerning knock on benefits of labelling for identifying weed problems] *"If you're turning labelling into record keeping, which is only another version of labelling, then in actual fact it would help because knowing what you did when and looking back next year you ought to be able to say.* (Participant B)

The broader political context was also seen to influence issues such as the economic outcomes:

Participant D: *"If it is better and more productive then economics should be good.*
Interviewer: *"Are you confident that it will be cheaper and better in economic terms?*
Participant D: *"That's where the subjective comes in Yes if the big six don't, – having got all the seeds and start charging premium prices for them – so it's a yes and a no.*

There were also particular differences in technical scoring on some issues such as the safety of organic food and the environmental impact of herbicides which highlighted where more technical information would be relevant to the debate:

For the organic, in theory they're supposed to be very safe but because they're not checked they could be very unsafe. And from the regulations... other regulations don't apply to them, so they don't score highly on safety. (Participant F)

The actual scores that were attributed in this exercise should not be considered definitive. Many of the criteria are complex and would need further disaggregation to score with more confidence. However, the people involved in the exercise do have considerable expertise in many relevant areas and collectively represent a wide range of pertinent technical perspectives. The general patterns in their scoring should provide, at the very least, a pointer to the broad character of the technical issues at stake.

Therefore, whilst bearing in mind the importance of context and the limited nature of this exercise, it is possible to draw out some general themes from the overall patterns in the actual scores themselves for the six basic options (ie: those for which an array of comparable scores were provided).

Environment

All participants score organic (or, in the case of Participant L, IPM) most highly under environmental criteria. All but one participant (F, a government adviser) score IPM higher than GM options under environmental criteria.

Under no viewpoint do the different regulatory contexts for the GM options significantly affect their relative environmental performance. Of the GM options, the 'voluntary controls' regime tended to be scored equal best or marginally higher in environmental terms under all but one viewpoint (Participant K, from the agriculture and food industry).

Under only one viewpoint (Participant F, a government adviser) were GM options assessed to perform significantly better in environmental terms than does conventional intensive agriculture.

Agriculture

The pattern displayed by the scoring under agricultural criteria is quite volatile, with four of the six basic options scoring most highly under one viewpoint or another: organic (Participants A and C), IPM (Participant J), conventional (Participants F and I) and GM with monitoring (Participants E and K). Likewise, all options score lowest or joint lowest under one viewpoint or another.

Health

The pattern of health scores is generally similar to those under environment, but are more variable in that, under two viewpoints GM options are regarded as performing better in health terms than do conventional crops (Participants E and F, both government safety advisers), with a third viewpoint (Participant K, from the agriculture and food industry) holding all non-organic options to be equally superior in health terms to organic cultivation.

Under no viewpoint do the different regulatory contexts for the GM options significantly affect their relative health performance. By contrast with environment, however, there is a slight tendency for the 'labelling' regime to score most highly among the GM options in health terms. This arose because some participants thought labelling would facilitate the early identification of any adverse effects and allow action to be taken.

A striking feature of this picture is that the most favourable assessments of the non-GM options under health criteria are significantly *less favourable* than the most favourable assessments of the GM options. Participant F (a government adviser) identifies a strong health advantage for GM over non-GM options on the grounds that there is greater regulatory oversight of GM foods.

Economics

Under economic criteria the organic option performs relatively poorly (scoring lowest under the viewpoints of Participants A, B, F, J and K, representing a variety of categories of participant). Only one participant (C, an independent scientist) differs (and strongly) by rating the organic option highest under a broadly-defined set of economic criteria.

Participants are evenly divided as to whether GM options as a whole display economic advantages or disadvantages. Participants B, F and K (associated with government and industry perspectives) see advantages; Participants A, C, D and J (associated with NGO and independent scientific perspectives) see disadvantages compared with conventional crops. Interestingly, all but two participants (A and L) rate the voluntary controls option as the worst or joint worst of the GM options under economic criteria. The discussion that took place on this point during scoring indicates that this was largely because it was judged that voluntary controls would restrict the areas of growth and thereby limit any economic benefits.

Society

Only five participants formulated scores under social criteria and the pattern in the scores assigned is similar in its volatility to the picture under agricultural criteria. In general, the GM options tend to score relatively low under the social criteria and the non-GM (especially organic and IPM) relatively high.

Other

Scoring data for the 'other issues' criteria is available for only three participants (C, D and Participant L). As a result of this restricted empirical base, generalisations over 'other criteria' are of very little value.

4.5 Uncertainties

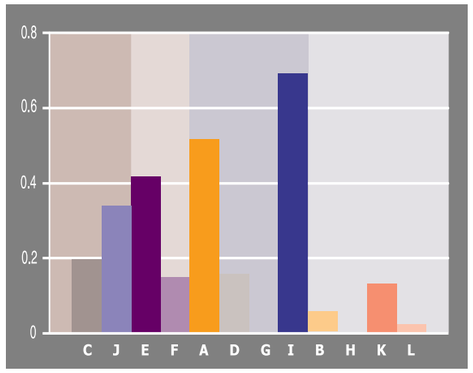
There was a significant difference in the degree to which uncertainty is expressed in the scores assigned by different participants. Indeed, there is a factor of ten difference between the extremes (when expressed as ratios to the mid-range values taken by scores under each individual criterion). The uncertainties expressed under each grouping of criteria by each participant is displayed in Figure 4. There is an evident tendency for participants from the agricultural and food industry (participants B, K and L) to fall among those with relatively lower levels of uncertainty across all the different criteria groupings.

Figure 5 displays the uncertainties expressed by each participant, broken down by each of the six criteria groupings. In general, the greatest uncertainties are expressed in the scoring of environmental performance, and (where they are assessed) the least with 'other issues'. Overall, agricultural, health and social issues are evenly ordered between these in terms of decreasing uncertainty in scoring. However, environmental, agricultural and health issues are all subject to the greatest uncertainty under one viewpoint or another.

Perhaps not surprisingly, significantly greater uncertainties are generally associated with the GM options than with the non-GM options (Figure 6). However, the appraisals of several individual participants display a significantly different pattern, with

Figure 4 The uncertainties expressed for each issue according to each participant

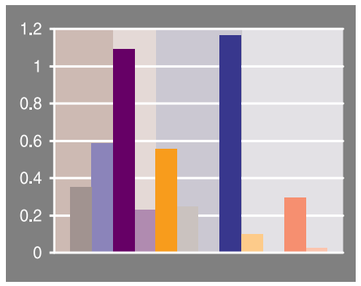
Overall



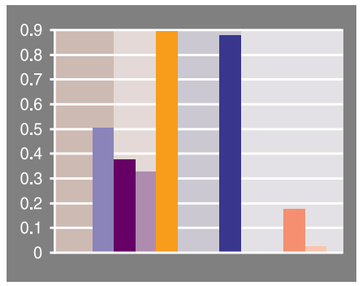
KEY TO PARTICIPANTS

- C
- J
- E
- F
- A
- D
- G
- I
- B
- H
- K
- L

Environment



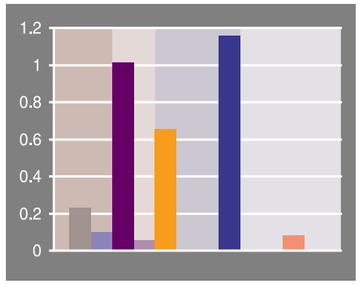
Agriculture



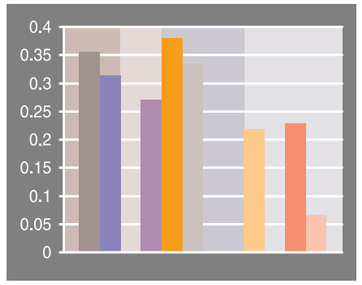
KEY TO GROUPINGS

- Academic scientists
- Government safety advisors
- Religious & public interest groups
- Agriculture & food industry

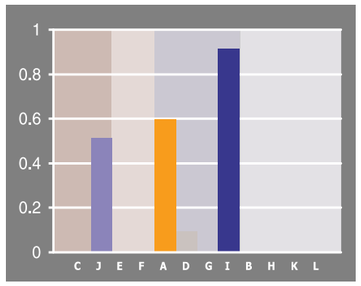
Health



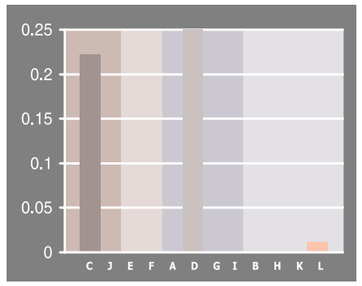
Economics



Society



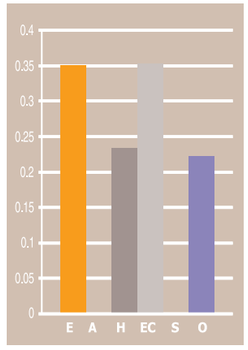
Other



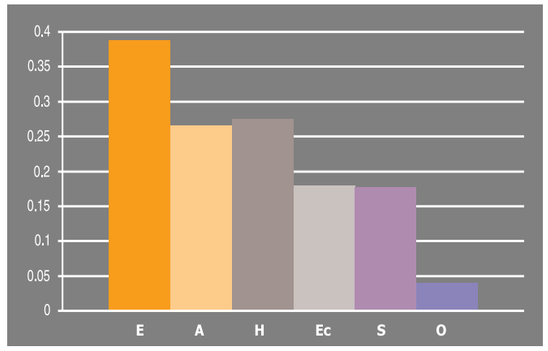
The degree of uncertainty expressed by each participant in scoring each group of criteria (environment, agriculture, health, economics, society, other) under pessimistic and optimistic assumptions. The uncertainty is expressed as a ratio to the mid-range between pessimistic and optimistic ranks. Each participant is represented by a colour coded bar. The average uncertainties across all groups of criteria are also shown separately.

Figure 5 The uncertainties expressed by each participant according to issue

C



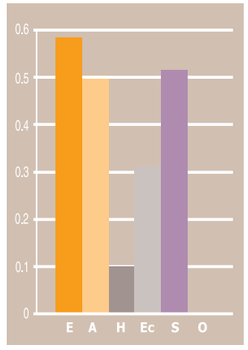
OVERALL



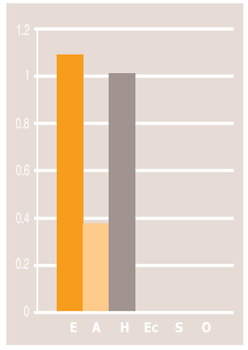
KEY TO ISSUES

- **E**nvironment
- **A**griculture
- **H**ealth
- **Ec**onomics
- **S**ociety
- **O**ther

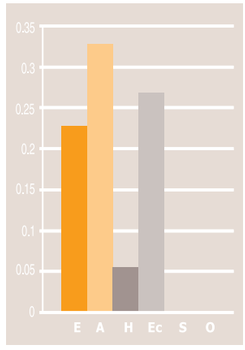
J



E



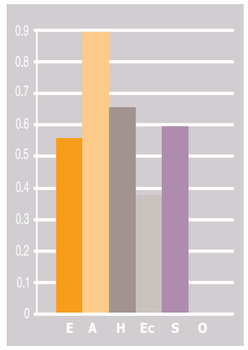
F



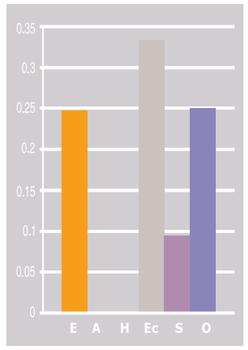
KEY TO GROUPINGS

- Academic scientists
- Government safety advisors
- Religious & public interest groups
- Agriculture & food industry

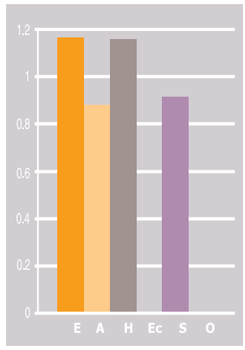
A



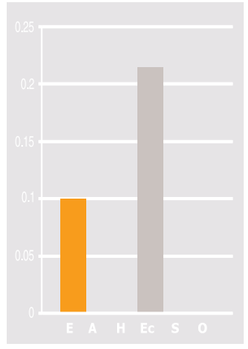
D



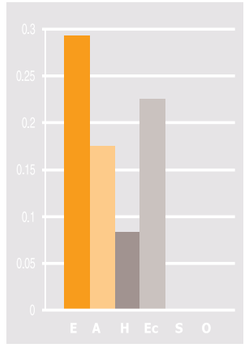
I



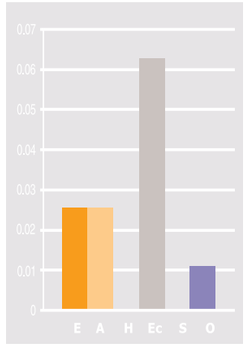
B



K



L



For each individual identified according to their code letter, a bar chart shows the uncertainty they expressed in scoring each of the isssues - environment, agriculture, health, economics, society, other and overall. The uncertainty is expressed as a ratio to the mid-range.

both organic and conventional farming subject to the greatest uncertainties under certain viewpoints.

4.6 Weightings

The assigning of numerical weightings to reflect the relative importance of different appraisal criteria is perhaps the most complex and potentially problematic aspect of an MCM analysis (cf: Section 4.1). Nevertheless, the overall picture displayed in Figure 7 has been validated in two separate iterations of consideration by participants (once during the interview and once for those responding to the sensitivity test).

Of the ten participants who felt fully able to assign weightings, eight identified at least one criterion in at least four of the six broad groupings of criteria (environment, agriculture, health, economics, society and 'other'). The averages of the weightings assigned to five of the six overall groupings are broadly comparable (ie: within factor two - 'agricultural practice' is the outlier). Of the six broad groupings of criteria, five are dominant under at least one perspective or another (the exception being the 'other criteria' group). This provides some further confirmation beyond the approval expressed by participants that grouping criteria in this way provides a relatively robust structure for thinking about the different types of concern.

The fact that five of the six broad groupings of criteria are dominant under at least one perspective or another reveals the magnitude of the differences in the perspectives taken by different participants. This observation is underscored by the fact that each of the broad groupings of criteria are entirely omitted under at least one perspective.

- No environmental criteria are weighted above zero by Participant F (a government adviser).
- No agricultural criteria are applied by Participants B, D or weighted by Participant F (including a government adviser, an NGO and an industry person).
- No health criteria are applied by Participant B (an industry person).
- No economic criteria are applied by Participants E or F (both government advisers). Participant I (from an NGO) rolls economics into one social criterion.
- No social criteria are applied by Participants B, E, F or K (drawn from government and industry sectors).
- No 'other' criteria (largely ethics) are applied by Participants A, B, E, F, H, I, J or K (drawn from virtually all categories of participant).

A special case in the assigning of weightings was Participant L who identified three criteria under which performance was not subject to trade-offs with that under other criteria, but which served rather as 'hurdles', which each option would have to pass if they were judged to be admissible as options. The three tests were 'regulatory approval', 'ethical acceptability' and 'commercial viability' (in terms of corporate strategy). With these hurdles passed, the scoring differences for the admissible options under the remaining nine of Participant L's criteria (falling under all six general groupings of criteria) were all weighted equally.

4.7 Rankings

Figure 8 displays the overall rankings for each of the six basic options under the perspective of each of the ten participants whose numerical 'inputs' permitted the derivation of multi-criteria performance rankings. It displays the outcome under both pessimistic and optimistic scoring. The axes are scaled in order to clarify *the differences in ranking orders* (rather than the absolute values taken by the ranks) under pessimistic and optimistic assumptions.

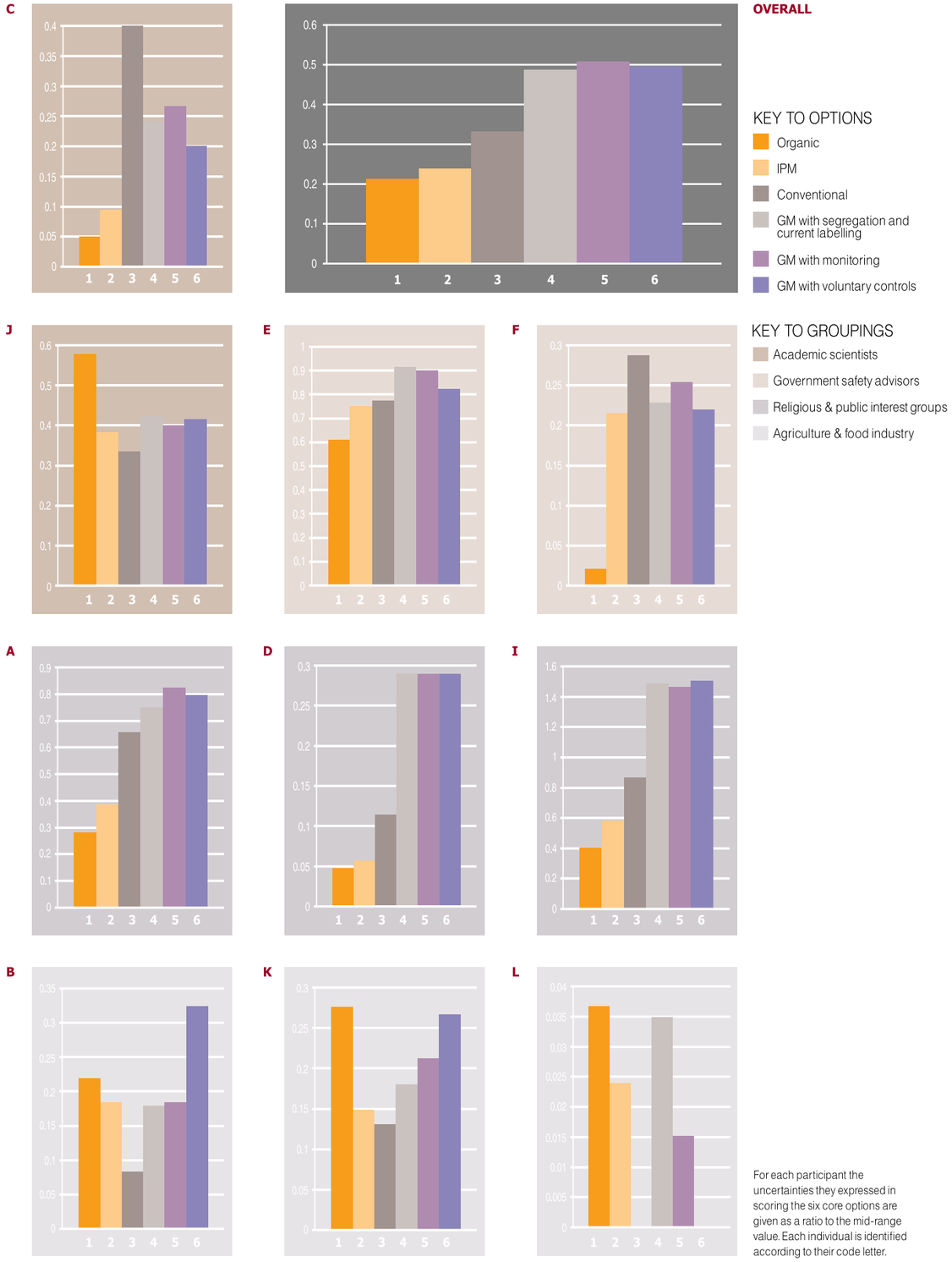
Several features emerge:

- the viewpoints taken by the different participants result in very different ranking orders across the six basic options.
- whatever the overall rankings, a distinction can be drawn between the pattern displayed by the three non-GM options (options 1-3) and the three GM options (options 4 - 6). However, there are important exceptions to this generalisation: the GM / non-GM dichotomy breaks down in some cases, with variation *within* these groups exceeding the differences *between* them.
- though there are cases where the differences between pessimistic and optimistic scoring are highly significant in the overall rankings, for the most part the differences between options are more pronounced than those between the high and low positions in the ranking orders taken by individual options.

Each of the individual basic options is found to perform worst and, for that matter, most are found to perform best - under the viewpoint of at least one or other participant. For instance:

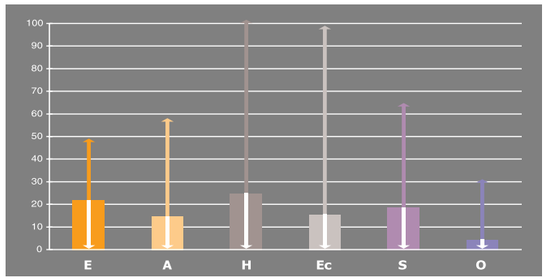
- Organic performs best according to participants A, C, D and I (and for an optimistic J and a pessimistic K) and worst according to participants B and (jointly) F.

Figure 6 The uncertainties expressed by participants according to the six basic options



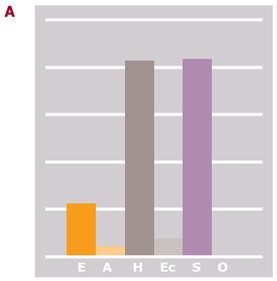
For each participant the uncertainties they expressed in scoring the six core options are given as a ratio to the mid-range value. Each individual is identified according to their code letter.

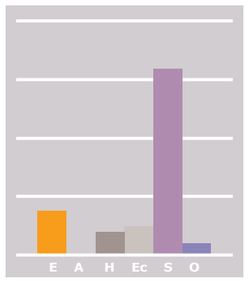
Figure 7 How each participant weighted each issue

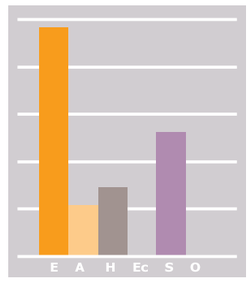


Average

Arrows show range from maximum to minimum value (%)

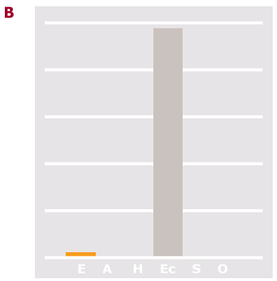


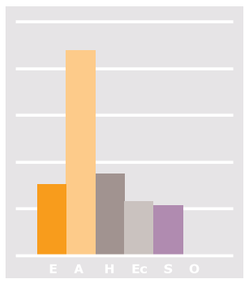


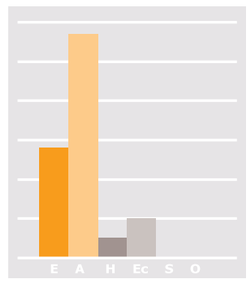


KEY TO ISSUES

- **E**nvironment
- **A**griculture
- **H**ealth
- **Ec**onomics
- **S**ociety
- **O**ther

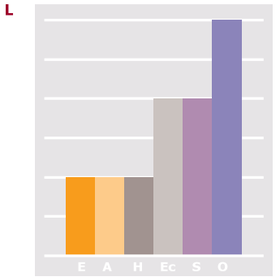


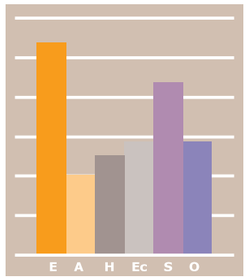


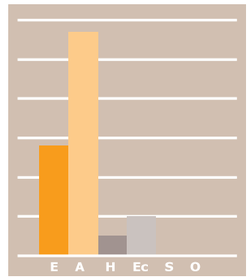


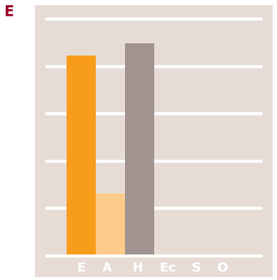
KEY TO GROUPINGS

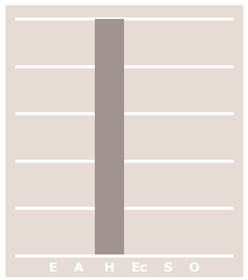
- Academic scientists
- Government safety advisors
- Religious & public interest groups
- Agriculture & food industry





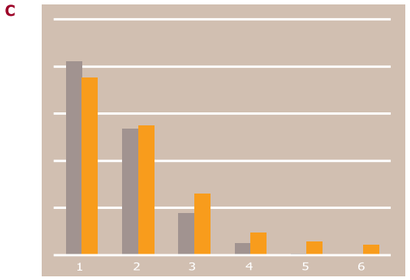


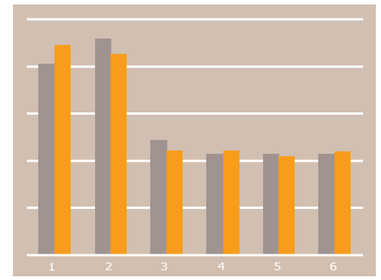




For each participant, identified by their code letter, a bar chart shows the relative weighting placed on each issue - environment, agriculture, health, economy, society, other (according to the colour code above). Also shown is the overall average of all participants in the top bar chart where the range is shown by the vertical arrows. The weighting scheme for Participant H is constructed from a simple ranking order. The scheme for Participant L omits 'threshold criteria' (see main text).

Figure 8 Final ranks for basic options by participant



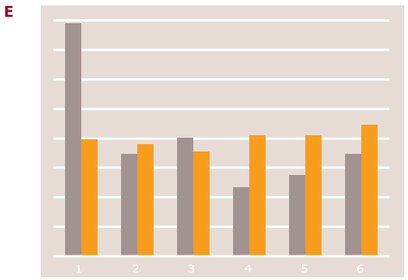


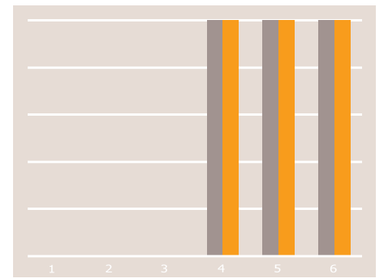
KEY TO UNCERTAINTY

- Pessimistic scores
- Optimistic scores

KEY TO GROUPINGS

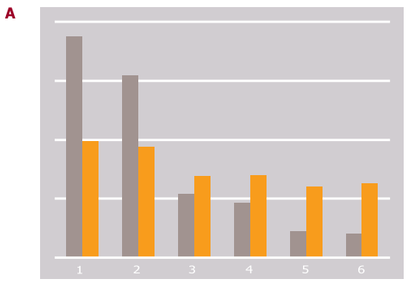
- Academic scientists
- Government safety advisors
- Religious & public interest groups
- Agriculture & food industry

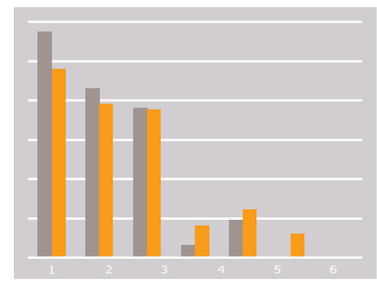


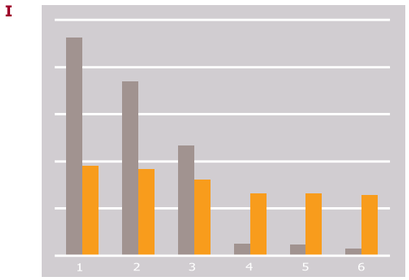


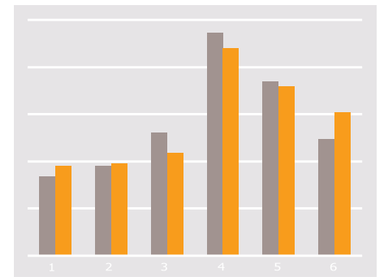
KEY TO BASIC OPTIONS

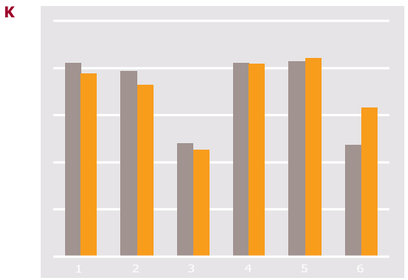
1 organic
2 IPM
3 conventional
4 GM with segregation and current labelling
5 GM with monitoring
6 GM with voluntary controls

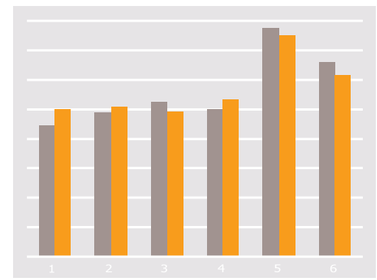












Each bar chart shows the final rankings for the six basic options obtained by each individual (identified by their code letter). The bars show the relative performance of the options according to an arbitrary linear scale. The pessimistic and optimistic bars display the uncertainty range in scoring that was expressed by each participant. The six core options are described in Table 4.

- IPM performs best for a pessimistic J and (jointly) worst according to participant F.
- Conventional cultivation performs worst according to participants K and (jointly) F.
- GM with segregation and labelling is found to perform best by participant B and (jointly) F, but (jointly) worst according to the pessimistic perspectives of J, I and L.
- GM with monitoring is found to perform best by participant L, an optimistic K (and jointly by F) and worst by participant J.
- GM with voluntary controls is found to perform best by an optimistic participant E and (jointly) by F and worst by participants A, C, D, I and (jointly) pessimistic J.

An option is assigned an overall rank of zero in this MCM procedure only when it scores lowest under *all* weighted criteria (both pessimistic and optimistic). Aside from the case of Participant F (who places a zero weighting on all but one criterion), this is the case only for option 6 (GM under voluntary controls) and occurs with two participants (C and D, an academic scientist and an NGO representative). The fact that zero rankings do not occur more often in this exercise is an illustration of a general willingness to score generally disfavoured options relatively highly under at least some criteria (cf: Section 5.2 on Strategic Behaviour).

Sensitivity Analysis: uncertainties in scoring

Figure 3 displays the overall rankings obtained by all the participants for their entire ranges of options under both optimistic and pessimistic scoring assumptions. Again, the absolute values taken by the ranks have been scaled in order to highlight the differences between the rank orderings under optimistic and pessimistic cases. These differences reflect the impacts of the more technical aspects of uncertainty (as distinct from issues of divergent interests, framing and values). The differences between rank orderings obtained under optimistic and pessimistic approaches to scoring under each individual perspective are evidently generally rather small compared to the differences that prevail between the perspectives themselves.

It is only relatively rarely that these overall uncertainties have any significant effect on the final ranking orders. According to the perspective of participant E, for instance, GM cultivation with voluntary controls moves from being the best-performing option under optimistic scores to being significantly worse than organic cultivation under pessimistic scores. Likewise, for participant J, a shift between optimistic and pessimistic assumptions alters the ordering (at the top end of the rankings) of organic production and IPM. In the vast majority of cases, however, the rankings of the different options remain unaffected by the uncertainties captured in the 'pessimistic' and 'optimistic' scoring conventions. The implications of this are that it is not the technical dimensions of uncertainty which are crucial, but rather more intangible qualitative aspects concerning the divergent interests, values and framing assumptions adopted by different participants.

Of course, the potential influence of technical uncertainties on rank orderings would be significantly larger if pessimistic and optimistic attitudes were adopted on a case-by-case basis with respect to *different* options or criteria rather than across the board as here. However, the fact that scoring is conducted in this exercise by the individual participants themselves already includes account of individual criteria- or option-specific framing assumptions. Further exploration of the importance of differentiated attitudes to uncertainty would rest on detailed examination of these assumptions and might be an interesting topic for further research.

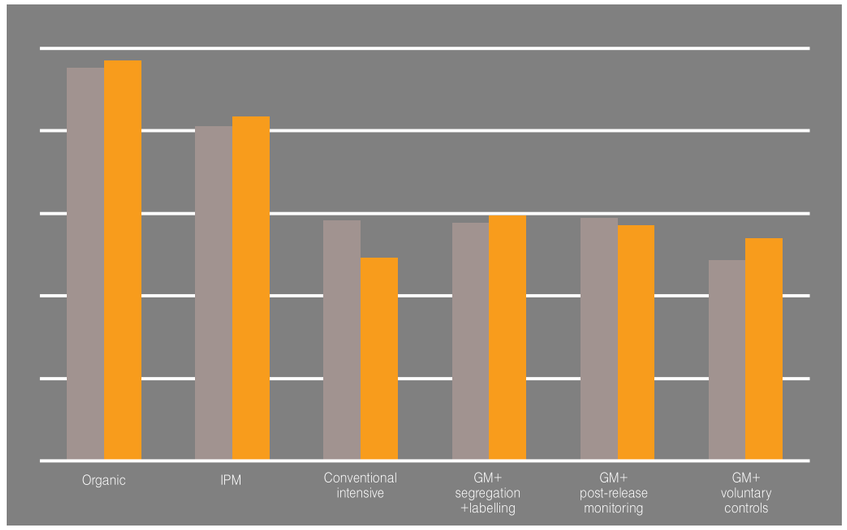
Sensitivity Analysis: importance weightings

A full table of the sensitivities of option rankings to changes in criteria weightings under each of the different perspectives is reproduced in Annex 2. It is remarkable that a ninefold variation in criteria weightings (factor three up and down from the base case) has such a relatively small impact (typically less than five percent) on the overall pattern in the rankings, only occasionally swapping the positions of options which are ranked closely together.

Nevertheless, there are a few examples where even threefold increases or decreases in weighting values yield apparently significant impacts on the final rankings. A factor three *reduction* in the weighting on environmental criteria under the perspective of participant E, for instance, changes the position of the organic option from being the most favourable to one of significantly lower performance than the GM options. Likewise, the same is true under this perspective for a threefold *increase* in the value of the weighting on health criteria. Similarly, under the perspective of participant K, the ranking of option 8 (a mixture of organic, conventional and conventional techniques) moves from a joint second position in the ranking orders (after organic cultivation) to being the best-performing option either under a threefold *decrease* in the weighting on agricultural criteria or under a threefold *increase* in the weighting on health criteria. The importance of these latter changes should not be exaggerated, however, because they take place against a background of rather close proximities in the rankings of the options concerned.

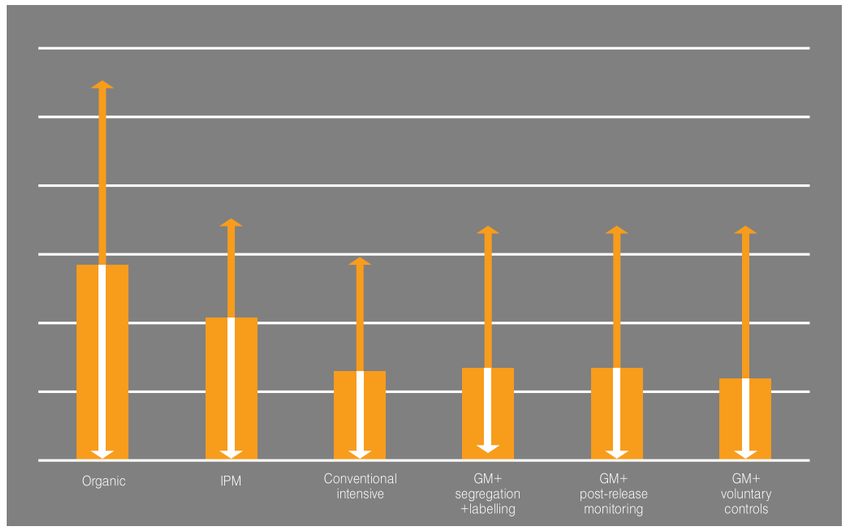
Overall it is clear that the final rankings obtained by the different participants and the broad associated

Figure 9 Average overall rankings for basic options



The average overall rankings for the basic options if the results from all participants are combined and equal weighting is given to each perspective. The orange and brown bars display the differences between the ranking orders over the six basic options obtained (respectively) under pessimistic and optimistic scoring assumptions. They thus provide an idea of the impact of technical uncertainties on ranking orders. (Since the numerical values for the ranks are on an arbitrary linear scale, the axes for the pessimistic and optimistic scores have been scaled separately to allow easy comparison of ranking orders).

Figure 10 Variability in overall rankings for core options



The bar chart shows the variability in overall rankings for the basic options using an arbitrary linear scale of relative performance. The bars show the average rankings. The vertical arrows show the maximum and minimum values obtained under the perspectives of different participants. Pessimistic and optimistic scores have been averaged.

patterns in their similarities and differences noted here are rather robust features of this exercise. They appear not to be volatile consequences of the weightings schemes alone, but rather are the result of a range of different aspects of the perspectives taken by the participants, including the choice of criteria and the frames of reference adopted in the assigning of performance scores.

Some Key Features of the Rankings

In the light of the relative robustness of the rankings to the sensitivity analysis concerning both technical uncertainties and importance weightings, some key features of the rankings can be identified with some confidence.

First, the GM options perform clearly best overall *only* under the perspectives of three of the participants associated with government or industry bodies (participants B, F and L). Under five academic scientific and public interest group perspectives (participants A, C, D, I and J), GM options perform generally worse than non-GM (especially organic and IPM) options.

Second, under the perspectives of two participants associated with government or industry bodies (E and K), the position is more equivocal, with non-GM options (notably organic) performing better under certain conditions.

Third, the voluntary controls regime performs generally indifferently or worst among the regulatory strategies for GM crops under the perspectives of *both* industry and public interest group participants alike. Only under the perspective of a government adviser (participant E) do voluntary controls appear unequivocally preferable to other GM regulatory strategies (under optimistic scoring) with the other government adviser neutral on this point (participant F).

Finally, it is evident that the conventional intensive cultivation option (option 3) tends to perform generally rather poorly under *all* perspectives, either with respect to the GM options or with respect to the organic and IPM options (depending on the perspective). Under no perspective is conventional intensive cultivation identified as the single best option. This finding is particularly interesting, given that this option represents the *status quo*.

A 'Bottom Line' Result?

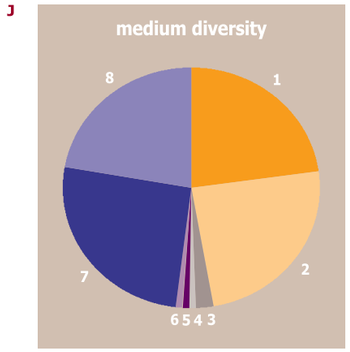
The 'bottom line' result obtained under most approaches to appraisal (such as cost-benefit or risk analysis) is a final ordering of options in sequence of their overall performance according to the particular assumptions and methodologies adopted in analysis. Where the perspectives taken by different participants are in some way aggregated in analysis, multi-criteria approaches are also often employed to yield a 'bottom line' result of this sort. By assigning equal weight to the viewpoints expressed by each participant and by assuming that the perspectives adopted are broadly commensurable with each other, such a procedure is also possible even in a simple 'linear additive weighting' pilot MCM exercise such as that employed here.

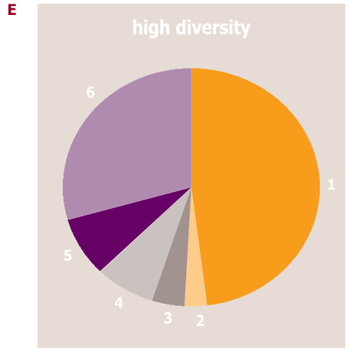
Despite concerns over factors such as the comparability of performance data, the commensurability of different types of benefit or burden, the consistency of framing assumptions and appraisal methods and so on it is often said that - irrespective of the difficulties - policy making *requires* the production of 'bottom line' results in appraisal, with corresponding pressures and responsibilities placed on analysts. The question might thus legitimately be raised as to what would be the aggregate ranking order for the six basic options under consideration here, if equal weighting were assigned to the perspective of each of the ten participants for whom rankings can be generated?

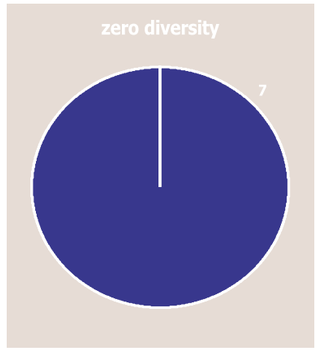
Having normalised and averaged the rankings for the six basic options under each viewpoint, the 'bottom-line' result for this exercise is displayed in figure 9. Organic cultivation is found to perform best on average, IPM next, with conventional agriculture and the three GM options all ranked similarly overall. Being averaged out in such a procedure, the variabilities between optimistic and pessimistic scoring conventions exert only minimal influence on this overall picture of ranking.

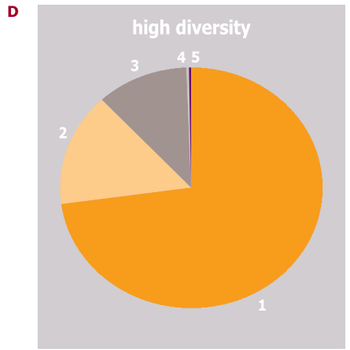
However, the confounding effect of the intrinsic uncertainty and indeterminacy in appraisal is shown in figure 10. Here, the big picture is one of enormous variability, with the overall rank of each individual option highly sensitive to the particular assumptions made under the viewpoints of the different individual participants. The extreme minimum and maximum rankings assigned to the different basic options under different perspectives are displayed as vertical lines in Figure 10 (together with the overall average or optimistic and pessimistic rankings as histogram bars). This shows that the overlaps and contrasts between the rankings of different options and the factors which drive these discrepancies are far more important than the differences between the average 'bottom line' option rankings. It is the systematic exploration of these factors which constitutes the real focus and contribution of a MCM approach such as the present pilot exercise and shows what a limited picture a 'bottom-line' analysis gives.

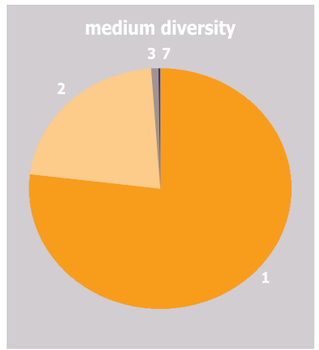
Figure 11 Pie charts showing diverse mixes of options favoured by each

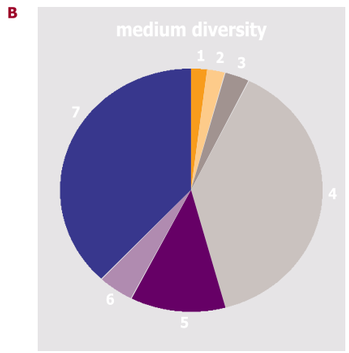


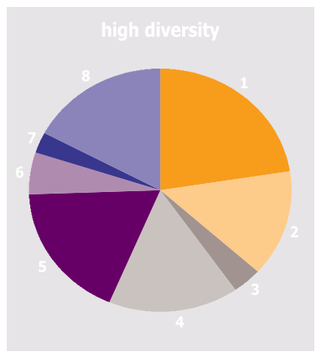












KEY TO OPTIONS

- Option 1
- Option 2
- Option 3
- Option 4
- Option 5
- Option 6
- Option 7
- Option 8

KEY TO GROUPINGS

- Academic scientists
- Government safety advisors
- Religious & public interest groups
- Agriculture & food industry

KEY TO OPTIONS
BASIC OPTIONS

1 organic
2 IPM
3 conventional
4 GM with segregation and current labelling
5 GM with monitoring
6 GM with voluntary controls

ADDITIONAL OPTIONS
(see Table 4)

I 7 GM with segregation, labelling, monitoring and other restrictions
8 GM with full assessment of impacts and need

B 7 GM with 'quality'

K 7 conventional and organic without GM
8 mix of GM, conventional and organic

J 7 GM in regulated organic system
8 GM in IPM system

F 7 GM with segregation, labelling and monitoring

Participants were asked their opinion concerning diversity among options. The results for each participant are shown as a pie chart displaying the proportional importance of each option in their favoured mix of options. The six basic options 1-6 are colour coded. The additional options defined by each participant are briefly described above. Full descriptions are given in Table 4.

4.8 Diversity

Figure 11 shows a set of pie charts representing the diverse mixes of options for the production of oilseed rape which were identified as the "more acceptable scenario" by those seven participants (B, D, E, F, I, J and K) who responded on this question. This group was drawn fairly evenly from government (E and F), industry (B and J), academic science (J) and public interest (D and I) backgrounds, rendering possible the making of some tentative observations concerning the practical role of diversity in an MCM appraisal of this sort.

The pie charts in figure11 represent, in approximate terms, the relative importance of each individual option in the overall mix of approaches to the cultivation of oilseed rape pursued in the UK taken as a whole. The measure of 'importance' here, relates to some straightforward 'output' metric such as share of production or share of land in cultivation. The details are not important because the important point in this pilot exercise is simply to introduce and explore in broad brush terms the implications of a relatively novel concept in appraisal - the devotion of explicit and systematic attention to the possibility of deliberately fostering some degree of diversity among choice options as a way of mitigating serious uncertainties and accommodating divergent value judgements.

Based on the procedure described in Section 3.6 and Annex 4, participants were invited to select one of four mixes of options constructed on the basis of their own final rankings across all their options, but including some degree of diversity. The 'zero diversity' scenario was constructed 100% from the best-performing option. The 'low', 'medium' and 'high' diversity scenarios represented the inclusion of progressively greater degrees of diversity, meaning that successively less well-performing options (under each individual perspective) began to be included in the mix of options.

Although figure 11 shows quite profound variability, a number of quite interesting features are evident. Only one of the seven participants (F a government adviser) indicated a preference for no diversity whatsoever. The remaining six participants (drawn from a wide variety of constituencies) all expressed an interest in either 'medium' or 'high' diversity. No participant selected a 'low diversity' scenario. This, together with the comments made by individual participants in their feedback on this stage of the exercise, indicates that there exists considerable empathy (at least among these seven participants) with the rationale for considering at least some degree of deliberate diversification among the better-performing options in order to hedge against intractable uncertainties and accommodate divergent social interests and value judgements.

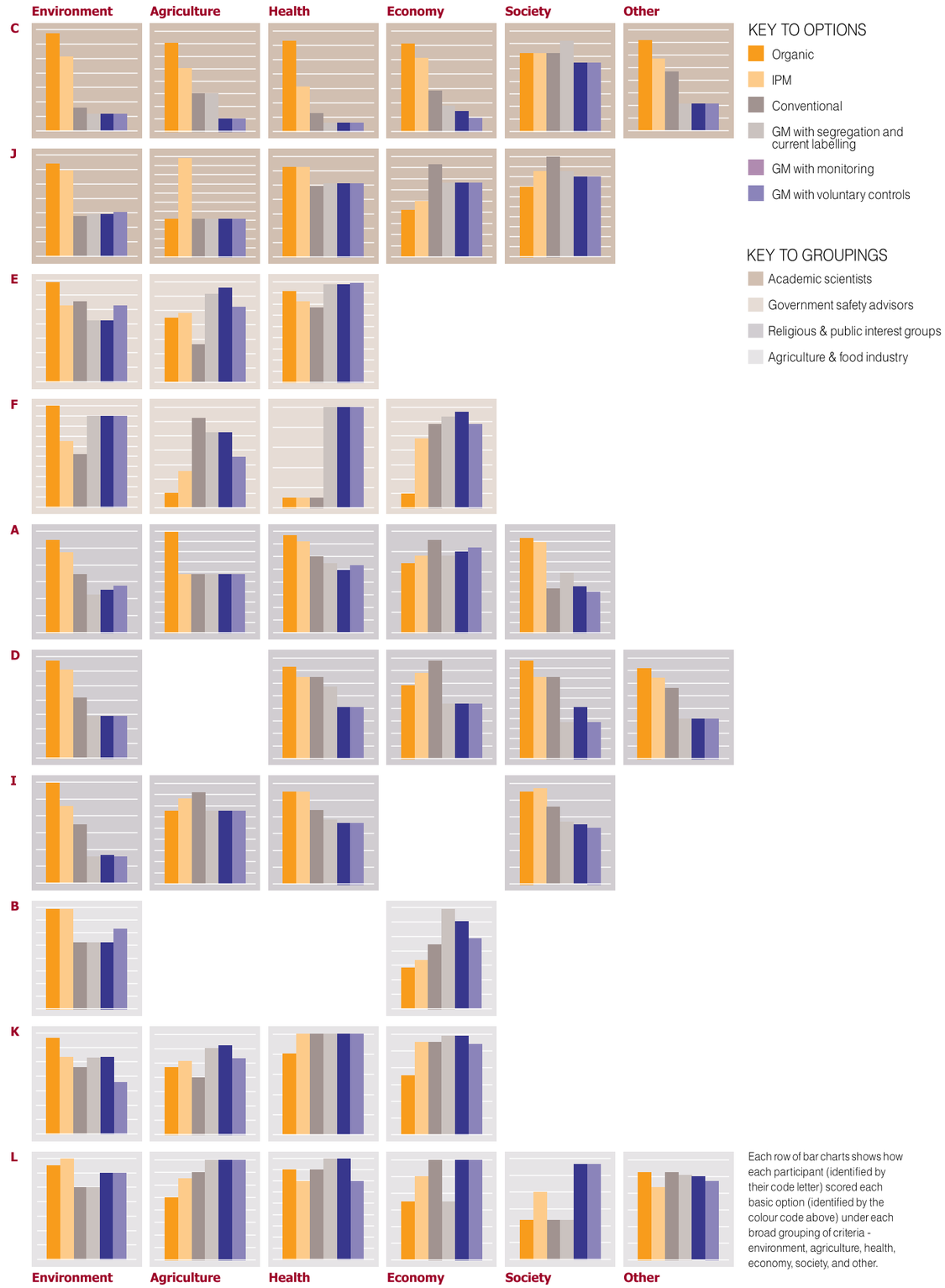
Among the six respondents who placed a non-zero weighting on diversity, it is striking that the two participants from NGOs (D and I) chose 'acceptable scenarios' displaying markedly less diversity than the remaining four participants (participants B, E, J and K). Assigning 'medium' and 'high' weightings to diversity, the positions taken by the two NGO participants are apparently no less well-disposed in principle towards diversity than are those of the other four respondents. It may be, therefore, that the relatively lower diversity of the scenarios favoured by the NGO respondents is simply a reflection of the greater differences between the final rankings taken by the various options under their appraisals. The greater the difference between the rankings of better and worse options, the higher the weighting that must be placed upon diversity in order to yield a given diversity of mix.

However, it is remarkable that the mixes of options identified by the two NGO participants whilst displaying a degree of diversity involving conventional agriculture, integrated pest management and organic cultivation both effectively *exclude all GM options*. This effective exclusion of GM options from these mixes is at one level simply a reflection of the relatively low rankings achieved by these options under these perspectives. However, one participant (D), commented that in the chosen mix of options small experiments with GM crops would not be ruled out under that perspective. The other (I) commented that the generally favourable view taken towards diversity under that perspective was qualified by recognition of the potential irreversibility of the adverse interactions between GM options and other cultivation strategies. Under such a view, the effects of GM strategies militate against diversity. Where appraisal (such as orthodox risk assessment) concentrates on the evaluation of individual options on a case by case basis, this kind of system-level consideration can easily remain neglected.

A further notable feature is that (with the exception of an industry participant (B) who envisages a contribution of only some two per cent) *all* participants envisage what would under present circumstances be considered very substantially increased contributions by organic cultivation methods.

A final striking observation is the consistently low importance under all perspectives assumed by conventional intensive techniques for the production of oilseed rape. What amounts to the current *status quo* is assigned at most only a few percentage points under views drawn from government advisers, industry and interest groups alike. This seems to underscore the observations already

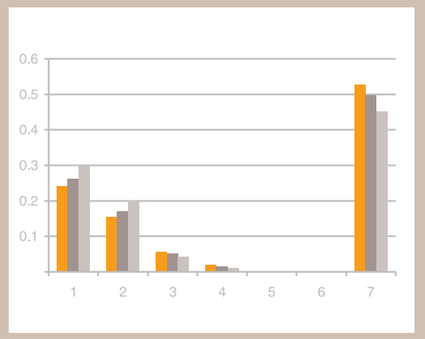
Figure 12 The participants' scores for each basic option



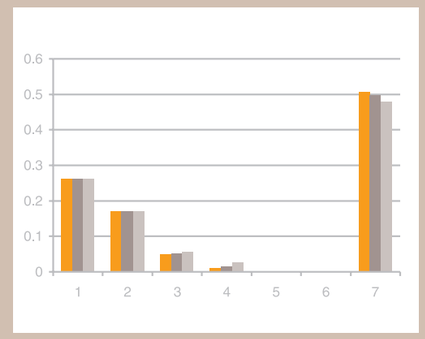
made with respect to the rankings themselves regarding the evident pervasive dissatisfaction with current techniques in comparison with alternative possibilities. Of course, the *particular* alternatives which are favoured under different perspectives are polarised between, on the one hand, integrated pest management and organic cultivation and, on the other, various GM strategies.

Participant C

Academic scientists



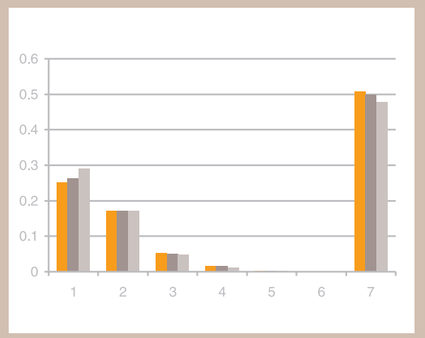
Environment



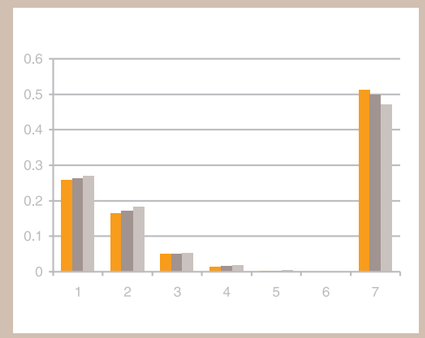
Agriculture

KEY TO SENSITIVITY TEST

- Reduce issue weighting by factor 3
- Base case weighting
- Increase issue weighting by factor 3



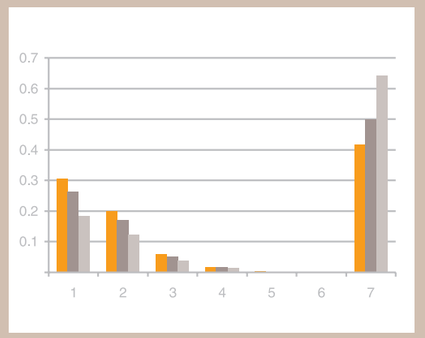
Health



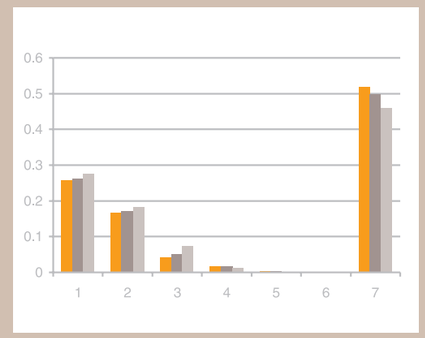
Economics

KEY TO OPTIONS

1 Organic
2 IPM
3 Conventional
4 GM with segregation and current labelling
5 GM with monitoring
6 GM with voluntary controls
7 complete public control



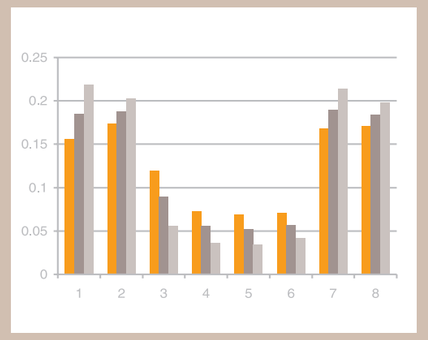
Society



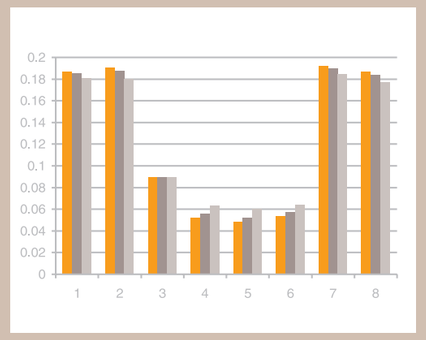
Other

Participant J

Academic scientists



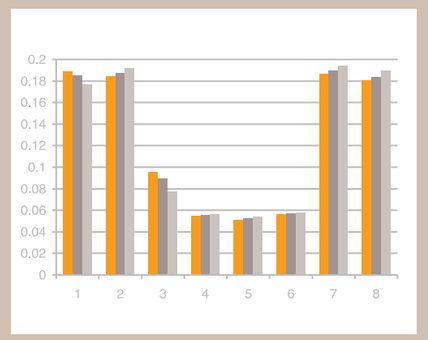
Environment



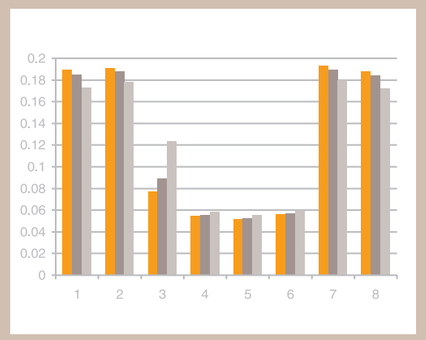
Agriculture

KEY TO SENSITIVITY TEST

- Reduce issue weighting by factor 3
- Base case weighting
- Increase issue weighting by factor 3



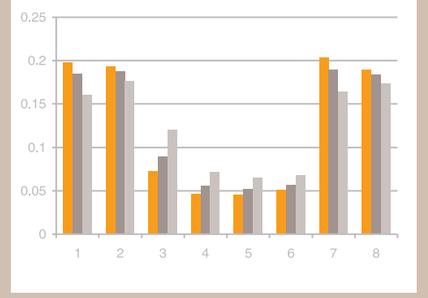
Health



Economics

KEY TO OPTIONS

1 Organic
2 IPM
3 Conventional
4 GM with segregation and current labelling
5 GM with monitoring
6 GM with voluntary controls
7 GM in regulated organic system
8 GM in IPM system



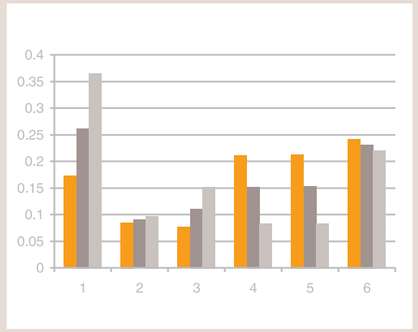
Society

No Criteria

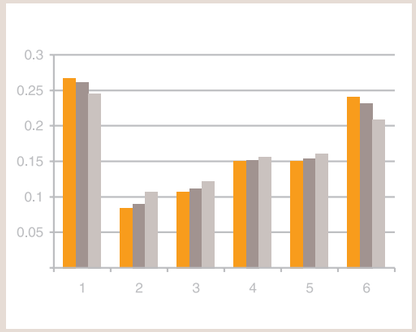
Other

Participant E

Government safety advisors



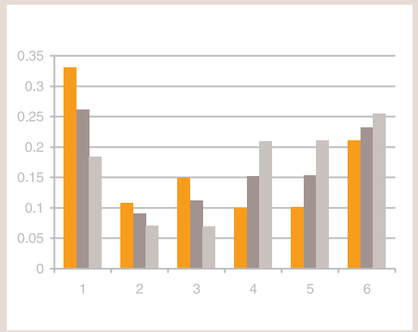
Environment



Agriculture

KEY TO SENSITIVITY TEST

- Reduce issue weighting by factor 3
- Base case weighting
- Increase issue weighting by factor 3



Health

No Criteria

Economics

KEY TO OPTIONS

1 Organic
2 IPM
3 Conventional
4 GM with segregation and current labelling
5 GM with monitoring
6 GM with voluntary controls

No Criteria

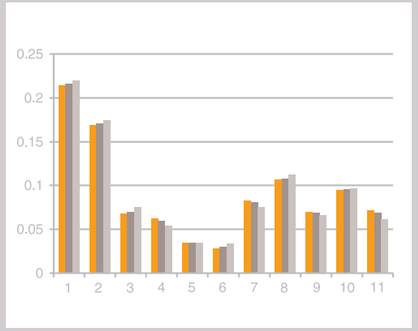
Society

No Criteria

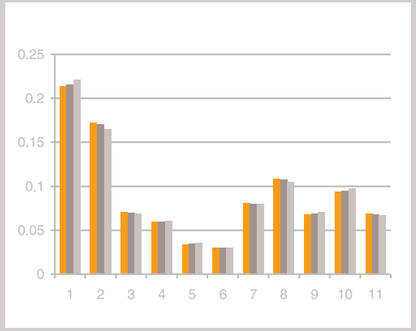
Other

Participant A

Religious and public interest groups



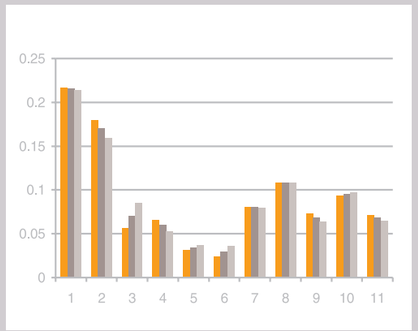
Environment



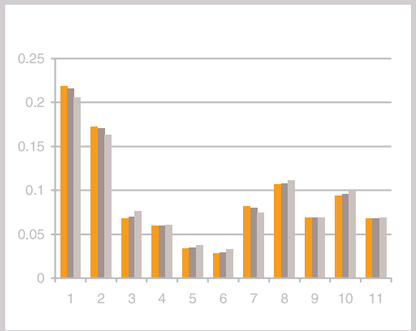
Agriculture

KEY TO SENSITIVITY TEST

- Reduce issue weighting by factor 3
- Base case weighting
- Increase issue weighting by factor 3



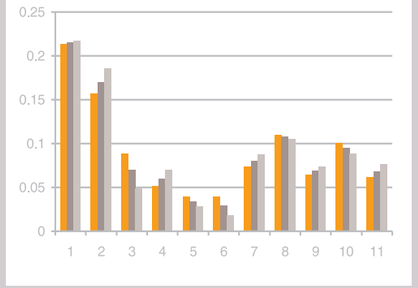
Health



Economics

KEY TO OPTIONS

1 Organic
2 IPM
3 Conventional
4 GM with segregation and current labelling
5 GM with monitoring
6 GM with voluntary controls
7 GM with labelling, monitoring and binding contracts
8 no GM commodity crops
9 GM in controlled sectors
10 GM only in USA
11 GM with legal safeguards



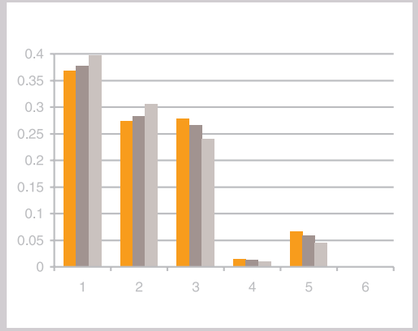
Society

No Criteria

Other

Participant D

Religious and public interest groups



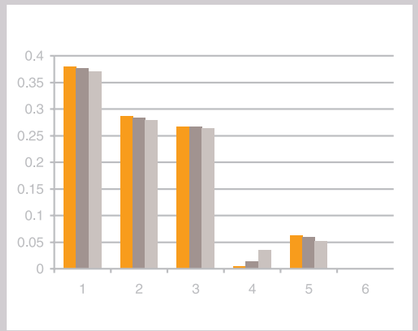
Environment

No Criteria

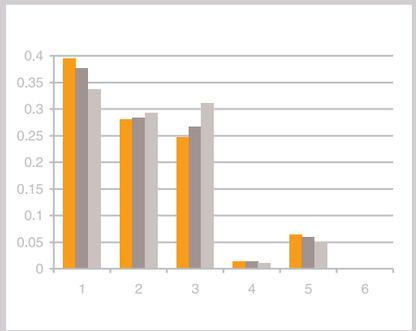
Agriculture

KEY TO SENSITIVITY TEST

- Reduce issue weighting by factor 3
- Base case weighting
- Increase issue weighting by factor 3



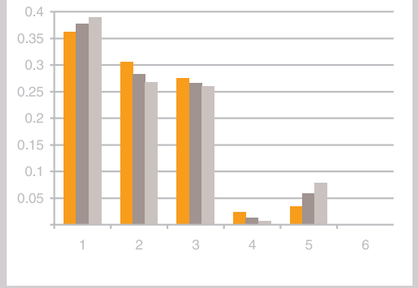
Health



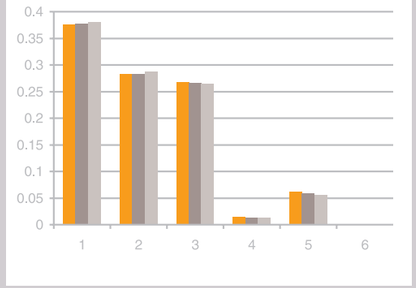
Economics

KEY TO OPTIONS

1 Organic
2 IPM
3 Conventional
4 GM with segregation and current labelling
5 GM with monitoring
6 GM with voluntary controls



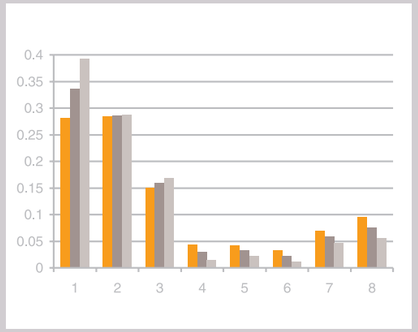
Society



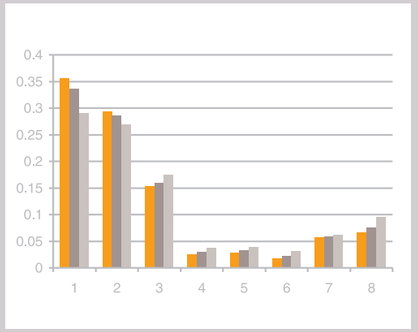
Other

Participant I

Religious and public interest groups



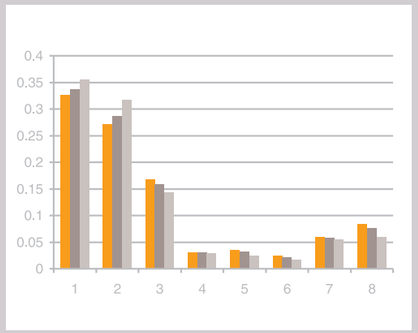
Environment



Agriculture

KEY TO SENSITIVITY TEST

- Reduce issue weighting by factor 3
- Base case weighting
- Increase issue weighting by factor 3



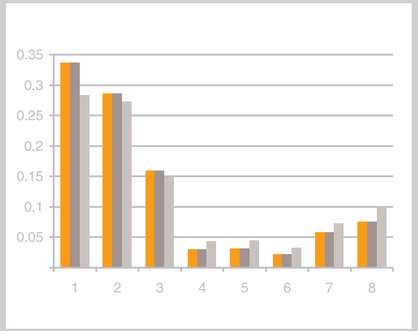
Health

No Criteria

Economics

KEY TO OPTIONS

1 Organic
2 IPM
3 Conventional
4 GM with segregation and current labelling
5 GM with monitoring
6 GM with voluntary controls
7 GM with segregation, labelling, monitoring and other restrictions
8 GM with full assessment of impacts and need



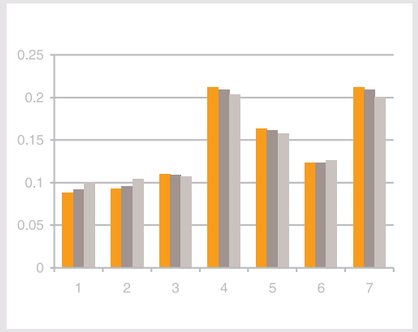
Society

No Criteria

Other

Participant B

Agriculture and food industry



Environment

No Criteria

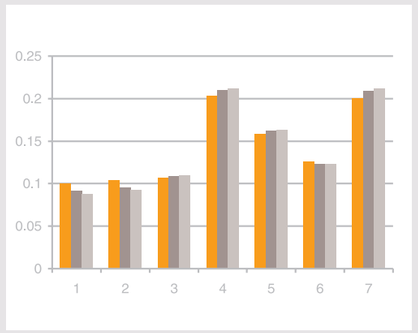
Agriculture

KEY TO SENSITIVITY TEST

- Reduce issue weighting by factor 3
- Base case weighting
- Increase issue weighting by factor 3

No Criteria

Health



Economics

KEY TO OPTIONS

1 Organic
2 IPM
3 Conventional
4 GM with segregation and current labelling
5 GM with monitoring
6 GM with voluntary controls
7 GM with 'quality'

No Criteria

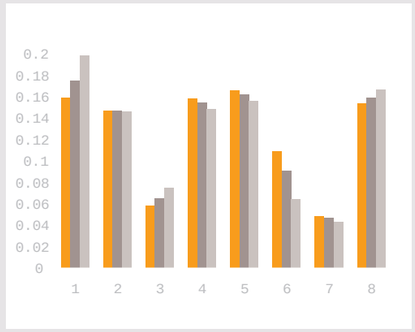
Society

No Criteria

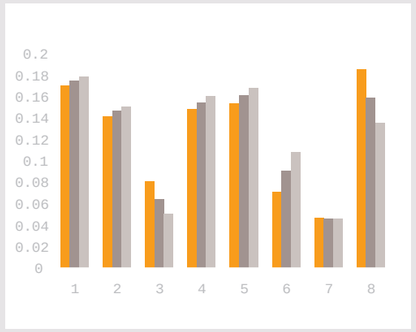
Other

Participant K

Agriculture and food industry



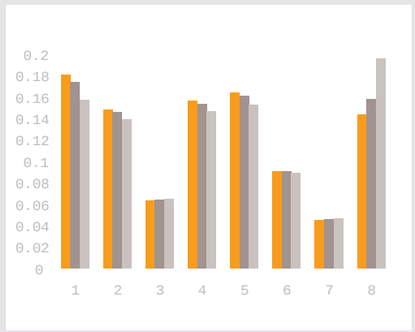
Environment



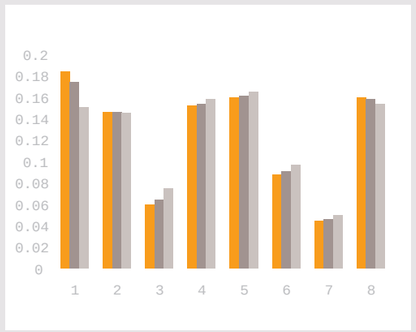
Agriculture

KEY TO SENSITIVITY TEST

- Reduce issue weighting by factor 3
- Base case weighting
- Increase issue weighting by factor 3



Health



Economics

KEY TO OPTIONS

1 Organic

2 IPM

3 Conventional

4 GM with segregation and current labelling

5 GM with monitoring

6 GM with voluntary controls

No Criteria

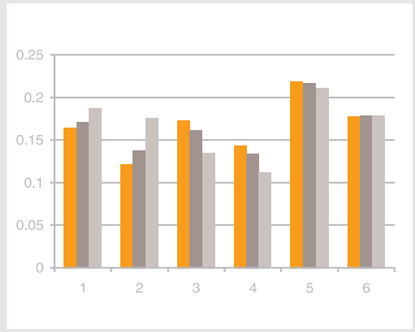
Society

No Criteria

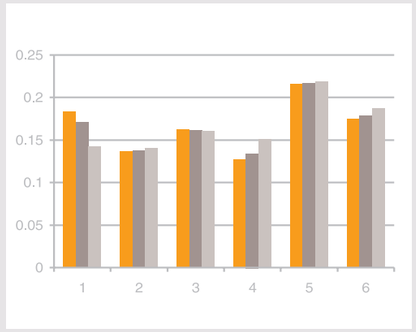
Other

Participant L

Agriculture and food industry



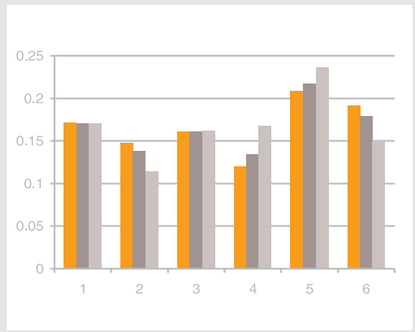
Environment



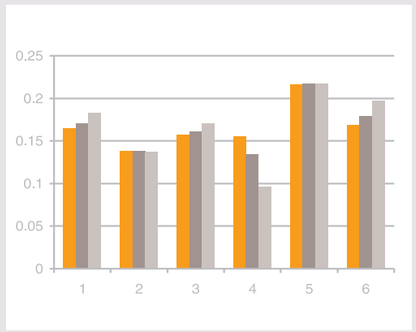
Agriculture

KEY TO SENSITIVITY TEST

- Reduce issue weighting by factor 3
- Base case weighting
- Increase issue weighting by factor 3



Health



Economics

KEY TO OPTIONS

1 Organic

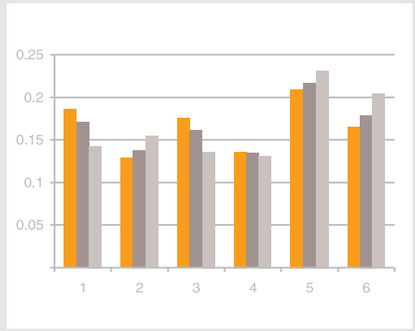
2 IPM

3 Conventional

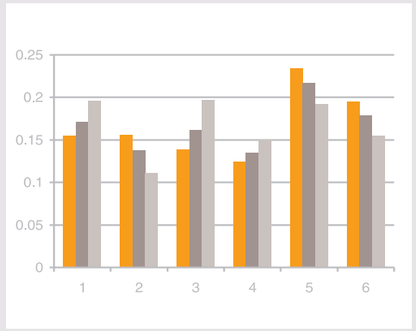
4 GM with segregation and current labelling

5 GM with monitoring

6 GM with voluntary controls



Society



Other

5. DISCUSSION

5.1 Engagement by Participants

Great efforts were made in this pilot exercise to adopt only the most straightforward of methods and to minimise the number and types of constraint placed on participants in the framing and treatment of the different issues. Nevertheless, engagement in such an MCM remains a demanding undertaking. Even the stripped-down, four-step procedure of option definition, criteria selection, performance assessment and criteria weighting remains a quite elaborate and challenging technical procedure for the uninitiated. However, the majority of participants (ten out of twelve) felt able fully to engage with the process, showing a strong intuitive grasp of what was involved and a readiness to deliberate in a disciplined and intensive fashion under an externally imposed framework and schedule. All participants responding to the interim analysis commented that they found the exercise worthwhile. No-one withdrew from the process.

The brevity of the personal interaction with participants in this pilot exercise (a single session of some two to three hours) may have been a disadvantage in that it may partly have been responsible for the incomplete engagement by two of the twelve participants (somewhat less than twenty per cent). This is significantly less time than is demanded of participants in most other full-fledged multi-criteria appraisals and compares with typical total commitments of several days in the context of a citizen's panel or consensus conference. A longer session might have assisted in the communication of the nature and intention of the exercise in these cases and so have fostered a greater understanding and empathy for what was required. Likewise, it is possible that lessons learned from this pilot exercise might contribute to improved communication on the part of the researchers.

However, there is also a sense in which the short duration of the process was also a positive feature, particularly in securing engagement from the particular key individuals involved in the present exercise. One of the concerns sometimes raised over participatory appraisal procedures is the time commitments required of the various parties including the analysts. All else being equal, the smaller the burden placed on participants the easier it is to elicit the involvement of influential specialists in the policy debate who are giving their time entirely voluntarily for a project outwith their own responsibilities and against the constraints of very busy schedules. Had the process been more demanding of their time, or involved the more complex scheduling requirements of multilateral deliberations, it is likely that there would have been greater problems of engagement.

It is also possible that at least some of the incomplete engagement in the present exercise is at least partly due to conceptual or evaluative difficulties with MCM on the part of individual participants. This is an important issue which does not receive the attention it deserves in the wider literature. Where incomplete engagement is a reflection of such intrinsic conceptual or evaluative difficulties, the less restrictive assumptions and less circumscribed scope of MCM should make it a less pronounced problem than with other (more elaborate) quantitative approaches to appraisal such as cost-benefit and risk analysis. The problems with engagement are notorious, for instance both in risk communication and contingent valuation and yet are often not fully declared. Where incomplete engagement does occur in MCM, however, it is important fully to acknowledge it and note the reasons for it. The relative flexibility and transparency of the MCM approach make this a relatively easy task.

It is interesting to consider the degree to which fundamental matters of principle are raised by the evident discomfort experienced by one or two participants in being asked to assign numerical weightings to reflect the relative importance of criteria. The various anti-utilitarian and wider philosophical concerns raised in the literature apply most acutely where MCM is employed to derive a final aggregated prescriptive 'solution', rather than a decision heuristic as in the present case. In any case, such concerns apply even more strongly with approaches such as risk assessment and cost-benefit analysis. However, if profound issues of principle are raised for some, then the particular principles concerned are evidently not shared by other participants. Some participants in the present exercise, for instance, not only felt comfortable in

weighting broad criteria of 'ethics', but were also content to assign numerical scores to reflect the performance of the individual options under such criteria. Ultimately the fundamental ethical implications of MCM will vary from case to case. As a heuristic, it certainly seems to have broader intuitive appeal than is typically the case with risk or cost-benefit analysis.

5.2 Interactions and Deliberation

It is important explicitly to highlight the nature of the interactive deliberation involved in this exercise. This raises two key issues. One (interactions with the researchers during the interviews) has been touched on in discussing the importance of context during scoring (Section 4.4). The other issue (interactions between participants) remains a relatively undeveloped feature of this pilot study.

The formal role of the researchers in framing and guiding the business of option definition, criteria definition, scoring and weighting was restricted simply to communication of the rationale for the methodology itself and to the formulation of the six 'basic options'. The procedures pursued in these areas have already been described in some detail (see especially Section 3.5). However there remains the question of informal, unintended and tacit influences arising through the dynamics of the personal interactions during the interview itself. The principal consideration here is that the researchers were careful to adopt a quite tightly disciplined approach, restricting their involvement as far as possible to responses to questions over methodology, with unsolicited interventions kept to a minimum and then only in the form of open-ended questions. Such questions arose, for instance, with regard to the elucidation of the specific reasons for differences in scores between options under individual criteria and were intended primarily to ensure the correct interpretations and recording of factors such as criteria definition.

For the most part, participants displayed a pronounced degree of self-sufficiency in the defining of options, the setting of criteria, and the assigning of scores and weights. Here, it is likely that the nature of this particular group of participants was important – all being specialists in the field with well-established and strongly-asserted positions independent of those of the researchers. If the participants had been lay members of the public, then it is likely that a more pro-active role would have been required of the interviewers, correspondingly increasing the scope for tacit influences. Whatever the reasons, it is significant that no participant at any stage raised the conduct of the interviewers as an issue. The fact that no participant wished to alter the positions arrived at during the interview itself, despite invitations to do this, also tends to confirm a sense (at least on the part of participants) that tacit influences by interviewers are a relatively minor issue. All things considered, any unintended influences on the results of this analysis are far more likely to relate to the relatively rigid structure imposed on the interview process by the MCM methodology itself than by personal interactions with interviewers.

The other significant issue of interactive deliberation (that concerning interactions between participants) is potentially more important. A crucial feature of the present pilot exercise is the restriction of the analysis to the exploration of positions taken by *individuals* in a relatively isolated methodological setting. Although (as discussed in Section 3.7) a final meeting was held, in the event this fulfilled the role more of quality control for the interpretation of results than of substantively influencing the character of the findings. In many ways, this relatively individualistic approach was dictated by the provisional nature of the present pilot exercise, and by the difficulties in reconciling the busy professional schedules of the various participants. As it turned out (and despite strenuous efforts on the part of the researchers) it was only possible to involve half the participants in the final meeting. A more intensive process of multilateral deliberation would therefore have required considerably greater lead times and resources.

This raises the question of the substantive importance of the lack of deliberation *between* participants in the interpretation of the results of this study. Might the outcome have been different if there had been greater opportunities for the mutual articulation, exploration and confrontation of different positions? This is an interesting and important issue which can only be resolved by further research. All that can be noted at this stage is that care should be taken not to make assumptions about the likely effect of such a process. Whilst experience in some instances is that extended interpersonal deliberation can achieve a degree of convergence and increased coherence between perspectives, it is also possible that such encounters can, under other circumstances, engender greater polarisation and entrenchment. It must be remembered that the present exercise was conducted against the backdrop of antagonistic positions in a high profile controversy with very large political and economic stakes for many of the parties engaged in this exercise. Either way, the potential for contingency, idiosyncracy and path-dependency in any particular discursive process is always likely to remain a crucial issue significantly compounding the complexities raised in considering interactions with the interviewer alone.

5.3 Strategic Behaviour

A further issue which like non-engagement is sometimes seen as a methodological difficulty (but which is not so categorised here), concerns the extent to which the inputs elicited from participants and their consequent results should actually be taken at face value. Even though it was made clear that the exercise is undertaken for heuristic rather than prescriptive reasons, it remains possible in principle that the expression of certain viewpoints might be subject (at least in part or unconsciously) to *strategic considerations* – the desire deliberately to influence the result in one way or another.

Although a feature of any analysis or deliberative process, concerns over the possibility of strategic behaviour are raised especially often of multi-criteria approaches, where the participatory element and the separate articulation of criteria choice and weightings are justified not technically but in terms of subjective values. Such issues are particularly pertinent in the present exercise where participants themselves (each being professionally engaged in the field in question) were asked to fulfil the role of an 'expert' (in scoring) as well as that of a 'citizen' (in criteria choice and weighting).

A variety of features of the results obtained in the present MCM might (depending on the perspective) be taken to raise questions in this regard. For instance, there are two instances where a certain option is held to perform worse than all others, irrespective of the criteria under which it is appraised (eg: participants C and D in figure 12). There are in some cases also some fairly strongly circumscribed constraints adopted in certain choices of criteria (eg: participants B and F in figure 7). The weighting scheme adopted in one case is such as entirely to exclude all but one of the criteria that have been identified as relevant (participant F). Other features of interest in this regard might be drawn from the contrasts in the scoring patterns of different individuals evident in figure 12, for instance with respect to the health merits of the different basic options (Participant F).

The point is, of course, that what may appear as expedient strategic assumptions under one view may be viewed as the manifestly reasonable and dispassionate framing of the issue under another. This is as true of the positions taken by the analysts themselves as it is of their subjects. In techniques such as risk and cost-benefit analysis, for instance, the choice of framing assumptions (concerning criteria, options, 'system boundaries' etc), parameter values (such as discount rates) and methodological conventions (eg: costing methods) may all readily have the effect of favouring one type of result over another and yet be defended in the context of a legitimate analytical perspective. Likewise, the influence of strategic framing arises from time to time with regard to the conduct of deliberative techniques such as consensus conferences and focus groups. There seems little reason to regard strategic behaviour as an issue which is specific to MCM. Indeed, it is arguable that the only effective way of addressing strategic behaviour in appraisal lies in the properties of transparency and verifiability which are argued elsewhere here to be better displayed in MCM than in many other approaches.

5.4 The Importance of Framing

An important feature of the MCM approach to appraisal is the explicit treatment given to different interests and subjective perspectives. The role in analysis of quantitative importance weightings is crucial in this respect. The results obtained by combining such weightings with performance scores may then either be aggregated to yield a final prescription, or systematically tested for sensitivity to divergent weightings as has been the emphasis in this exercise. Either way, the implicit assumption is often that all-important questions of divergent priorities and values can effectively be captured by the heuristic device of numerical weighting models.

The results obtained in this exercise raise quite serious questions over the validity (or at least sufficiency) of such assumptions. Despite the relative novelty and complexity of the procedure, participants were able to come to closure in their deliberation over weightings significantly more readily than they were over the scoring. Despite a number of opportunities and stimuli to rethink or change the weighting schemes on which they had settled in the interviews, participants seemed to be quite confident in the initial positions which they took on these weightings. In the sensitivity analysis, the general structure of the final rankings obtained by the participants remained remarkably stable over relatively large scale variations in weightings. For that matter, a similar general stability of rankings was displayed in the face of the inclusion of uncertainties by means of optimistic and pessimistic scores. And yet despite this apparent relative insensitivity to changes in the explicit parameters intended to model their differences, the disparities between the results obtained under the different perspectives remained quite profound.

In the present exercise at least, the crucial determinants of the differences between perspectives lie at least as much in choice of criteria and the qualitative framing assumptions adopted in the scoring of the different options under these criteria, as they do in the numerical values taken by the weightings or the

explicit uncertainty ranges. The practical considerations which govern these 'framing assumptions' are well documented in the critical literature on risk assessment where they exert no less an influence on the results, but all-too-often remain unexplored. Issues such as the 'system boundaries' of the options in question, the timescales over which appraisal is applied and the treatment of the passage of time, the feasibility, cost and acceptability of remedial or regulatory measures, the trajectories which the development of options are expected to undergo and the contingent influence of external events, all provide important dimensions for legitimate and defensible differences in the scoring of options.

It is a positive feature of an MCM approach that the explicit attention to divergent perspectives allows the better documentation of the practical importance of such framing assumptions. This information allows better understanding of the sometimes rather unexpected features in the performance of certain options. For instance, labelling may not only provide consumer choice but the potential to track and record effects and so this option performs relatively well under health criteria. In this way the broad scope of an MCM may throw up unexpected benefits or, equally, disbenefits of a particular option.

When these issues are drawn together, it seems on the basis of the present exercise, that there should be greater caution in assuming that differences of perspective can be *fully* captured in the quantitative weighting models of multi-criteria approaches in general. The basic structure evident in the option rankings seems to be governed at least as much by implicit factors in the choice of criteria and the scoring of options, as it is by divergences in importance weightings.

5.5 Relationships between Criteria

A final issue which is touched on intermittently in several places in this study and returned to in the context of methodology in Annex 3 concerns the degree to which the criteria formulated by each individual participant may be regarded for practical purposes as being mutually independent. In seeking to establish which performance characteristic is preferable in any given context, it is sometimes necessary to know first the status of some of the other performance characteristics. For instance (to take an often-cited example) before settling on a preference for red or white wine at a meal, it is necessary to know (at least for some people) whether the accompanying course will be meat or fish. This sort of issue can readily become quite complex, and can potentially arise under all approaches to appraisal which extend their attention to a variety of different factors, including cost-benefit and risk analysis. At root, however, the basic problem is quite simple.

First and foremost, it is necessary to be clear about what is *not* involved in considering this issue of criteria independence. It is not just a question over the existence of *correlations* between performance criteria. A tendency for poor environmental performance to be associated with poor health performance (for instance) need neither pose nor imply any necessary problems in terms of the independence of the associated criteria. The problem arises rather where a judgement over what constitutes good performance in one area *depends* on actual performance in another area.

Likewise, this issue of criteria independence is not the same thing as the imposition of performance thresholds as a prerequisite for the consideration of options. Under one perspective in the current exercise (that of participant L), for instance, compliance with regulatory requirements and consistency with corporate strategy were formulated as two criteria in their own right, in a sense which transcended the importance of all other criteria. In other words, the satisfaction of these performance thresholds was regarded as a precondition for the inclusion of an option in the appraisal. This approach to the prioritisation of criteria is well documented in the literature, where it is referred to as 'lexicographic ordering' the ordering of options as if they were words in a dictionary, with the performance characteristics as the letters. It is a very different approach to the analytical framework adopted in the present exercise under which all criteria are in principle traded off against each other. However, this type of relationship between criteria is quite readily dealt with simply by establishing the circumstances under which each individual option may be held to satisfy the threshold criteria, determining whether these circumstances hold and then applying the remaining criteria in the same fashion as for other participants. This was the approach taken in the present study to the two threshold criteria formulated by participant L (of Section 4.6).

With this clarification of two other forms of relationship between criteria, attention can return to the question of criteria independence. Perhaps the most likely indication of the intrusion of criteria dependencies in the strict sense during the present exercise might be expected during the scoring process. Any important dependencies between criteria would be likely to show up during the process of eliciting detailed justifications for the scoring rationales and contextual assumptions from each participant. Here, it can be reported that at no stage and for no participant did this emerge as a practical issue with respect to the criteria actually employed.

For those contextual factors which remain implicit, a further safeguard in the present exercise is provided by the explicitly *holistic* weighting procedure. Instead of arriving at criteria weightings through a process of pairwise trade-offs, the present approach was simply to elicit judgements over the relative importance of each criterion taken in the context of *all the others* (and bearing in mind the respective differences between best and worst performing options under each criterion). Along with a host of other contextual factors, any residual dependencies would feature as part of this deliberative process of judgement over weightings. In any case, the issue of criteria independence becomes just one among many other qualitative factors of context- dependency in framing of the kind which have already been discussed in the last section.

6. CONCLUSIONS

6.1 Validity

Before drawing conclusions from the results of this pilot MCM it is important to be clear about the conditions and qualifications to the validity that may be claimed. There are clearly limitations to this pilot exercise. Only twelve participants were involved, of whom only ten (a little over eighty per cent) completed the MCM procedure in its entirety. The participants were selected informally and cannot be seen to represent a statistically valid or otherwise representative sample. Beyond this, further more minor reservations may be expressed concerning the information available to participants in the scoring exercise, the amount of time available for the weighting of criteria, the relationship between the focus on oilseed rape and the general references to GM strategies in general, and the extent to which the final results of this study have been tested by in-depth discussions among *all* participants.

This said, there are many senses in which, even as a pilot exercise, the principal findings obtained in this study may (if interpreted carefully) confidently be regarded as robust. Although only twelve in number, the participants embodied an impressive array of expertises and institutional experiences and spanned a very wide range of the perspectives currently reflected in the debate over GM crops. All participants are professionally engaged in the issues associated with GM crops and food production at a level where it would not be surprising to find any of them serving on an official advisory committee of some form. The scoring process in particular was the subject of an impressive degree of careful deliberation on the part of all participants, with the researchers serving continually to challenge and document the consistency of assumptions. Beyond this, the present exercise is far less circumscribed in scope than are other typical appraisals in this field, lending a greater degree of completeness to the picture generated. Finally, the results obtained have been subject to a fairly intensive process of validation, both during the interviews themselves, through subsequent bilateral consultations and in discussions involving a fairly representative cross-section of one third of the participants.

To the extent that all appraisal is necessarily a collective undertaking on the part of a number of specialists, the present exercise is no different. Indeed, it may confidently be argued that the range of pertinent professional perspectives represented in the present exercise is significantly greater than that which is typical in an orthodox risk assessment. The degree to which uncertainties and discrepancies between the positions taken by different participants have been made explicit by the MCM methodology should not be mistaken for a lack of robustness on the part of the exercise as a whole. On the contrary, the tendency in conventional appraisal to exclude, evade and even deny differences of perspective should rather serve to render this more complete, transparent and systematic exercise all the more robust.

As long as care is taken not to extrapolate beyond the present results to generalisations concerning the balance of views among different constituencies or society as a whole then, as a heuristic exercise, the present pilot study may cautiously be regarded as a source of a number of quite interesting and relatively robust insights concerning the general structure and dynamics of the current debate over oilseed rape in particular and (with more care) over GM crops in general. Perhaps the most useful way of interpreting these results is as a potential guide to the design of further appraisal research in this field, which might usefully set out to build on these foundations and address some of the gaps, ambiguities and question marks which must necessarily be left in a pilot study of this sort.

6.2 Mapping the Debate

The main objective of the present pilot MCM exercise is to serve a heuristic, rather than a prescriptive, function. In this sense, the utility of the results lies as much in insights concerning the structure and dynamics of the current debate over GM crops as in the normative implications for agricultural strategies or regulatory policy. In this regard a number of conclusions may be drawn, some of which confirm findings made elsewhere, others of which suggest surprising and potentially significant challenges to certain received wisdoms.

First, there are other options which are thought to be viable and broadly comparable with the pursuit of GM strategies, at least with regard to oilseed rape. Combinations of organic and GM strategies were one of these and (under the perspectives in which they have been formulated) tended to perform relatively well in this exercise.

Second, it is clear that a very wide range of criteria are thought relevant to the evaluation of GM crops and alternative food production strategies, many of which are quite remote from the narrow scientific and health issues addressed in orthodox risk assessment. The implications here are returned to in a later section. For the moment, though, it is clear that an important group of criteria address issues not only of consumer choice but also of citizenship and wider questions of participation and agency. This picture echoes that currently emerging elsewhere in the literature. Analyses of the failings of present regulations point to exactly these two issues.[48,49] A recent workshop concluded that unless broader issues were included in the evaluation of GM foods, then the system will struggle to gain public support.[50] Another recent workshop also concluded that, in risk domains such as those to do with GM crops, a broader spectrum of knowledge was needed to inform the decision and public participation was crucial.[51] The issue of agency and ability of the consumer to make choices or influence decisions as a consumer has been highlighted in public attitudes research both in the UK and the Netherlands.[52] The results obtained here tend to confirm this broader picture.

Third, with regard to the performance of GM and non-GM options under health and environmental criteria, questions are raised concerning certain assumptions which might otherwise have been taken for granted. Although the differences in the pattern of option scores under these groups of criteria are quite strongly influenced by the viewpoint of a single participant, they underscore that it cannot be assumed that performance under health and environmental criteria will *necessarily* be well-correlated. In particular, it seems that there is less consensus over the health implications of GM crops than there is over their environmental performance.

Fourth, with regard to the perception and treatment of uncertainties, the implications of the present study are that the consistent adoption of 'optimistic' or 'pessimistic' approaches to the scoring of options does not generally affect the picture of overall performance as much as do differences in framing assumptions (concerning criteria choice, scoring and weightings). The differences between rankings obtained under optimistic and pessimistic scores are generally rather small compared to the differences between perspectives. The conclusion here, then, must be that it is not the technical dimensions of uncertainty which are the key issue, but rather more intangible qualitative aspects concerning the divergent interests, values and framing assumptions adopted by different participants.

Fifth, and also with regard to the matter of uncertainty, some confirmation is given here to a broad brush picture in which the greatest overall uncertainties are held to lie under environmental criteria and concerning GM options, with generally lower uncertainties tending to be seen by participants drawn from an industry background than by other participants. However, this broad impression overlies a richer texture of small-scale variability, with significant uncertainties also identified under different perspectives for *all* the broad groupings of agricultural, health and economic criteria and for conventional as well as organic production methods. The perception of uncertainty and variability is thus a highly complex and context-dependent factor casting doubt on automatic assumptions that the key uncertainties necessarily concern the environmental and health effects of GM crops. Under some perspectives, for instance, uncertainties over the large scale economic effects of a switch to organic farming present similarly profound issues of ignorance and precaution.

Sixth, with regard to the priority assigned to the different groups of criteria, the picture is perhaps not surprising. Perspectives drawn from the biotechnology industry and food supply chain are conspicuous in their relative under-emphasis of the social and/or environmental and safety considerations which are prominent under all other perspectives. The perspectives adopted by government advisers hold in common the distinctive characteristic of being at the same time relatively narrow in scope whilst emphasising environmental and safety considerations. For their part, the perspectives expressed by the non-industry participants (ie: the academic scientists, government advisers and religious and public interest groups) hold in common a markedly lower emphasis on economic or agricultural considerations.

48 Von Schomberg, 1998.
49 Wynne and Mayer, 1999.
50 Barling et al 1999
51 Confronting Risk: Finding new approaches to risk. Report of a workshop run by the Consumer's Association, Unilever and Sainsburys, 1998.
52 Hamstra, 1995.

Finally, with regard to the notions of overall performance yielded by the rankings obtained in this exercise under each perspective there are a few interesting conclusions. GM options perform best overall only under the perspectives of government or industry participants whilst they perform generally worse under the perspectives of academic scientific and public interest participants. However, even under certain government and industry perspectives, non-GM options including notably organic cultivation perform better under certain conditions. Perhaps most surprisingly, the voluntary controls regime performs worst or joint worst among the regulatory strategies for GM crops under the perspectives of *both* industry and public interest group participants alike.

6.3 Policy Implications for Agriculture

For a policy maker charged with making decisions over the regulation of a GM crop or a class of GM crops, what implications might be drawn from this pilot study? Although the conclusions drawn must necessarily be qualified as tentative, the results discussed here are no less robust (though for different reasons) than those typically yielded by comparable appraisal exercises in this field. Further investigation on a number of issues might change significantly the complexion of the results but the findings obtained in this study at the very least demonstrate the general form of the insights which might be thought likely to emerge from the more detailed, widespread and sustained use of MCM techniques in this area.

First, and perhaps most strikingly, it is clear in many ways that there is a generally quite favourable picture of the performance of organic systems of production. The superior environmental performance of organic techniques is a matter of consensus among participants. However a range of widely perceived broader benefits from organic strategies were evident not only from the overall rankings that emerged under a wide range of perspectives and the associated diversity analysis, but also from the choice of additional options for appraisal. Even those most positive about the technology consider that if GM crops could be included in an organic system then this might offer the 'best' option. This raises serious questions over the extent to which R&D strategies presently support such a progression towards organic and IPM systems or allow the more detailed evaluation of their feasibility and implementation.

Second (and almost irrespective of perspective), conventional intensive agriculture was seen to perform consistently poorly. The lesson from this may be that when evaluating GM or other developments in agriculture there may be some merit in going beyond the use of the conventional agriculture *status quo* as the 'yardstick' by which harm is evaluated. At present, for example, when deciding whether a GM crop will have an adverse effect on the environment, the UK's Advisory Committee on Releases to the Environment consider that an effect which is no greater than that caused by conventional systems cannot be considered an adverse effect.[53] The present results raise questions over how demanding a criterion this may be. It may be that regulatory appraisal of individual options would be more robust if they were carried out on the basis of comparison with a wider range of alternatives than just the present *status quo*.

Third, although there was evidence of support for controls on GM crops for a variety of reasons ranging from consumer choice, consumer confidence and the ability to track effects, there was some scepticism from a variety of perspectives over whether voluntary controls would be effective. Doubts were evident both with regard to the feasibility of such controls *in principle* once GM crops have been released and with regard to the confidence that may be placed in the actual observation of voluntary controls in practice. Most noticeably, the addition of options with a wide variety of post-commercialisation controls did not have a marked effect on the general performance of the GM options. That this result was sustained over such a disparate array of perspectives underscores questions over the confidence that may be placed in the effectiveness of such voluntary controls.

Finally, there was evidence from this exercise that considerable support may exist on all sides of the debate for the focusing of greater attention on the deliberate pursuit of a relatively *diverse* mix of agricultural strategies drawing on a number of the better-performing options, rather than on a single monolithic 'best' technological or policy option. This raises questions over the extent to which R&D and regulatory policy making should be geared towards active encouragement of a variety of techniques rather than assuming or emphasising a single particular trajectory. It also raises the issue of how to treat options which display characteristics which are seen to militate against diversity.

53 Von Schomberg, 1998.

6.4 Policy Implications for Regulatory Appraisal

The Validity of a 'Bottom Line' Result

Orthodox approaches to regulatory appraisal (such as risk and cost-benefit analysis) are routinely employed to arrive at discrete, apparently definitive and often highly precise results. These are then taken as an influential basis for subsequent deliberations over the 'safety' or 'acceptability' of individual options or more rarely the relative performance of a range of alternatives. Attention is more often directed at the simple values taken by such results than it is at the nature, context-dependency and defensibility of the crucial determining assumptions which led to the obtaining of these particular values rather than some others.

It has been shown in the present study that multi-criteria appraisal techniques may be employed in a similar fashion to arrive at a discrete set of prescriptive results. The justification for such an approach is no more questionable in principle with these approaches than it is in the fields of risk or cost-benefit analysis. However, the explicit attention given in MCM to the various crucial dimensions of appraisal (option choice, criteria choice and definition, the framing of performance scoring, the treatment of uncertainty, the weighting of importance) serves strongly to undermine the unqualified presentation of a particular prescriptive set of results. Rather than being seen as a disadvantage of MCM, such insights might rather be seen as crucial in the interpretation of the more opaque pictures yielded by orthodox appraisal. Conversely, where an MCM heuristic *does* admit qualified prescriptive conclusions, these may be regarded as correspondingly more robust than pronouncements made without any systematic attention to the volatility or idiosyncrasy of crucial determining assumptions.

Breadth of Scope

The wide spectrum of criteria that were thought relevant to the appraisal of GM oilseed rape under virtually all perspectives in this exercise raises serious questions over the scope of existing approaches to the regulatory appraisal of GM crops in the UK. Such concerns have already been widely voiced, for instance in recent statements by the Royal Society[54] and Royal Commission on Environmental Pollution.[55] Even where appraising performance in the relatively narrow terms of environmental and health impacts, issues were raised by a wide range of participants which are not exclusively concerned with the technical details of the method of production and so are presently entirely *excluded* from current approaches to regulatory appraisal. In the domain of environment, for instance, aesthetics and impacts on biodiversity are examples. With health, nutritional consequences were considered relevant by some. This picture is compounded in considering many of the social, economic and even agricultural criteria raised by participants from all sides of the debate, which are also routinely excluded from the procedures of regulatory appraisal for GM crops. In this light, the broadening of the scope of the regulatory appraisal process may be seen to offer an important way of improving the match with the wider debate, with corresponding implications for the fostering of trust and the reduction of polarised conflicts.

A similar point might be made with regard to the essentially comparative character of the present exercise involving consideration of a wide range of different options rather than the examination of individual options on a case-by-case basis under some absolute yardstick of performance (such as 'safety', 'risk' or 'cost'). There can be little doubt that the deliberation by participants over the conduct of scoring across a variety of options significantly enriched, extended and refined the exercise, by continually suggesting new factors or novel implications of established understandings. Likewise, the discipline imposed by the need continually to compare and contrast helped to elicit a better understanding under each perspective of the nature of the relative strengths and weaknesses of the different options which might have remained marginal in a stand-alone appraisal. Finally, of course, there is always the possibility that a 'satisficing' approach to regulatory appraisal seeking simply to establish the acceptability of a single option in isolation may all-too-easily lead to the neglect of alternative options which might otherwise have performed even better.

The Importance of Risk Characterisation

A related but distinct issue arises from recognition of the apparently relatively modest importance of the weighting process in determining the pattern observed in the MCM results. Although the numbers involved are too small usefully to bear statistical examination, it appears that the basic structure evident in the option rankings is governed at least as much by the choice of criteria and by the divergent framing assumptions adopted in scoring, as by divergences in weightings. Here, there may be

54 Royal Society, 1998

55 RCEP, 1998

important implications for the design and interpretation of multi-criteria techniques themselves, in that it seems that divergent knowledges, values, interests and other commitments are only partly reflected in the weighting schemes, being expressed also through all the other qualitative stages and elements in the appraisal process.

Beyond this, there are potentially significant wider conclusions which may be drawn with regard to the business of risk assessment as a whole. For instance, the above finding seems to reinforce the rationale for emphasising the importance to policy makers of the 'risk characterisation' process as voiced (for instance) in an influential recent report by the US National Research Council[56] and the EPA Commission.[57] Drawing a contrast between 'risk characterisation' and the more programmatic, quantitative aspects of risk assessment, the NRC observe that:

> *In addition to the biological and physical outcomes that are typically covered, decision makers and interested and affected parties often need to know about the significant economic costs and benefits of alternatives, secondary effects of hazard events, or the efficacy of alternative regulatory mechanisms* (page 29)

> *A risk characterisation will fail to be useful if the underlying analysis addresses questions and issues that are different from those of concern to the decision makers and affected parties* (page 29).

It is obvious that an appraisal process which excludes what are held by some constituencies to be important factors, may fail to secure the crucial property of public confidence. It follows from this that, by instilling a misleading impression of completeness, robustness or rigour, risk assessments based on such incomplete risk characterisation may leave regulators and business highly exposed to a subsequent backlash on the part of the excluded constituencies.

Likewise, it might be concluded from this that the addition of 'ethics' as a separable (and often final) 'bolt on' stage in the process of regulatory appraisal may also often prove inadequate and misleading. It is clear from the present exercise that values, interests and other commitments are all inextricably intertwined with the application of 'knowledge' in appraisal. Appraisal procedures which are predicated on the separation of these elements seem likely to fail.

The Potential of MCM

Overall, this exercise may be concluded to have demonstrated that MCM *does* offer a way of combining relatively technical and intrinsically subjective factors in appraisal in such a way as to display (at least to some extent) each of the properties outlined in Table 1, namely:

i) relative flexibility and breadth of scope,
ii) openness to divergent choices, values and framing assumptions,
iii) candour about uncertainties,
iv) a heuristic for 'mapping' (rather than prescribing) assumptions,
v) systematic discipline and rigour,
vi) transparency and verifiability under external review,
vii) accessibility to participation,
viii) feasibility and efficiency as part of a regulatory process.

The limitations which have been acknowledged and discussed in this report underscore that MCM cannot be regarded as a panacea for the complex and intractable challenges of risk assessment and technology appraisal. The complexity of the exercise and of the results means that MCM can certainly not be regarded as an everyday tool. It can only make sense, for instance, as part of a wider deliberative process of appraisal - a process within which it might be hoped that MCM may help contribute the key properties of systematic discipline, transparency and verifiability.

Finally, of course, it is clear from the plurality and relative open-endedness of these results that an MCM such as this *cannot* be seen as an 'analytical fix' for arriving at definitive 'right or wrong' answers over what constitutes the 'best' (ie: most 'reasonable' or most 'consensual') choice of option from the point of view of society as a whole. Far from being a difficulty, the lack of pretensions in this regard are a positive feature of the use of an MCM heuristic. For instance, this was evidently a condition for securing the trust and involvement of the unusually broad array of interests participating in the present exercise.

Here, the crucial point is that, while a technique such as MCM may be used to identify and explore the relative importance and interactions of issues such as option choices, framing assumptions, value judgements, uncertainties and technical evaluations, the final decision and its associated justifica-

56 NRC, 1996.

57 EPA, 1997

tions must remain at least to some extent intrinsically contingent and subjective. As a result, an important feature of an MCM approach is that it makes it explicit that the justification of final decisions must be as much in terms of political legitimacy and democratic accountability as in terms of 'sound science' or 'rational economics' in appraisal.

For a politician it may appear that an apparently simple, ostensibly precise, 'safe', 'unsafe' or 'safe enough' verdict may be more appealing than having to look at the more messy subjective factors in comparative appraisal. In reality, however, the opposite may be the case. Apparently simple conclusions of this sort are often rather poorly sustained by the real complexities of appraisal. They are widely contested and no longer serve the purpose of reassuring the public. Being able to justify decisions and show that all relevant criteria have been considered at some point in an evaluative process is likely to lead to more robust decision making. Practically speaking, it would not be necessary to repeat an entire MCM exercise for each individual GM crop, for example. Once the framework and general parameters have been set, the specifics pertinent to an individual decision might be added relatively easily. Thus used in the right situation, MCM could facilitate policy and decision making at many levels.

Further Work

From this pilot study it appears that an MCM technique, when used as a heuristic, could offer a potentially effective contribution to policy making and decision taking in many domains including biotechnology and agriculture. The main shortcoming of this pilot study, in terms of extrapolation and therefore practical usefulness, was its necessarily limited nature. A logical extension would be to expand the scope in three ways:

1. By providing for greater interaction and deliberation *between* participants.
2. To introduce a dimension of public participation by establishing a citizen's panel or panels to select additional options and criteria and assign weightings.
3. To use a variety of specialists agreed by the panel to score the criteria under the various options.

Because the language used and the issues raised by the present participants were those of the policy debate, it is particularly important that any subsequent exercises include wider publics both in order to identify any contrasts with the specialist arena and to confirm and enrich the 'map' of the overall GM debate. Panels could be selected on a regional basis, by age, sex or some other basis to bring different perspectives into the debate. This would serve both as a way of testing and verifying the present exercise as well as refining and rendering more robust the final picture.

ANNEX 3

A Technical note on the Multi-Criteria Mapping Methodology

The technique employed in this study is a 'heuristic multi-criteria mapping' exercise, using a 'linear additive weighting' approach.

The technique is *heuristic* because the principal aim is to explore the issues and come to a better understanding of the nature of the problem and the possible responses, rather than to make claim to the definitive determination of a single 'optimal' solution. It is a *mapping* exercise because the results are expressed systematically in terms of sensitivities prescriptive conclusions being drawn only conditionally, by reference to the clearly-defined perspectives taken by different participants. In both these respects, the approach differs from some other multi-criteria methods which use much more complex techniques in an attempt to identify a unique, determinate and 'objectively optimal' resolution of the various divergent perspectives.

The *linear additive weighting* aggregation model is based on the simple weighted average of option performance:

$$r_i = \Sigma_c \, s_{ic}.w_c \qquad [1]$$

In other words: the overall performance rank obtained for the i^{th} choice option (r_i) is the sum of the performance scores determined for that option under the c^{th} appraisal criterion (s_{ic}) each multiplied by the importance weighting on that criterion (w_c). The scores are normalised such that:

$$s_{ic} = (m_{ic} - m_{c,min}) / \Sigma \, (m_{c,max} - m_{c,min}) \qquad [2]$$

In other words: the performance score for the i^{th} choice option under the c^{th} appraisal criterion (s_{ic}) is the ratio of the difference between the performance measure determined for that option (m_{ic}) and that for the lowest-performing option ($m_{c,min}$) with the difference between the performance measures determined for the highest- ($m_{c,max}$) and lowest- ($m_{c,min}$) performing options under that criterion.

It will be immediately apparent to specialists in the field of multi-criteria appraisal that the method adopted here represents one of the simplest of all possible theoretically-valid approaches. As already discussed in the report, this simplicity represents a deliberate choice, reflecting the heuristic rather than prescriptive aims of the study and a concern not to allow the quantification procedure to obscure important qualitative features of the appraisal. In any case, for all their complexity, none of the many elaborate techniques developed in multi-criteria evaluation over the past four decades may claim fully and finally to have resolved the fundamental theoretical problems of social appraisal (such as the interpersonal comparison of utility and the formal impossibility of definitive social preference orderings).[58] It therefore remains an open question whether the loss of simplicity and transparency is worth the sometimes marginal improvement in fidelity.

For instance, no attempt was made in this exercise to aggregate and order the different criteria under a single over-arching *value tree*. Such an approach would have assumed the consistency of the performance measures applied by different participants under apparently similar criteria when, in fact, the differences between the ways of framing apparently similar criteria was a major qualitative finding of this study (Section 5.4). As discussed in Section 5.5, the formal *independence* of the criteria was established by checking for this property in the scoring exercise and further safeguarded by the holistic procedure employed in the assigning of criteria weightings under which many different 'framing' factors (including independence) were taken into account by participants themselves.

Likewise, the assigning of technical performance scores by the participants themselves is also a point of contrast with many multi-criteria studies, which often use a separate panel of specialists to deter-

58 Cf: Kelly (1978), MacKay (1980), (Collingridge, 1982), Bonner (1986), Bezembinder (1989)

mine a single set of scores under each criterion. The reason for the present approach is that the participants in this exercise are themselves professionals with respect to different aspects of the broad field in question. Indeed, the fact that such different scoring schemes were generated by participants in this exercise is itself evidence that a single set of specialist performance ratings cannot adequately address the complex considerations which arise in the scoring of options. The choice of specialist for an 'expert scoring panel' would thus clearly be a major variable and potential point of contention in its own right. In this regard, the establishing of alternative scoring scales by specialists from a variety of perspectives in an exercise such as this might offer a useful basis for the provision of scoring information for subsequent MCM studies involving non-specialist participants.

In similar vein, it is a feature of the present analysis that *all* scores are expressed as arbitrary rating scales, with *no use made at all* of physical metrics or established indicators. Again, this is an open reflection of the importance of divergent framing assumptions in scoring – a factor not avoided by the imposition of apparently precisely-defined indices. As is often displayed in risk and cost-benefit analysis, the use of a particular metric does not guarantee consistency in framing assumptions concerning such factors as discounting, system boundaries and the aggregation of micro-criteria. Here, the use of arbitrary scales allows the individual specialist participants to capture the entire range of what under their own perspective arise as the pertinent considerations under each criterion. Of course, the same would not be true of lay participants in an MCM.

A fourth point concerns weightings. The simple scalar weightings used in this exercise do not seek to model any non-linearities which there may be in the relationship between performance measures and subjective values. These factors are sometimes formally addressed, for instance, by *value functions*. However, such complex relationships are addressed in this exercise by the unconstrained and reflexive (with respect to rankings) character of the weighting procedure and by the crucial role of sensitivity analysis.

Also on weightings, it is a feature of this pilot study that the procedure involved the direct entry of weighting values by the participants, based on a 'holistic' appreciation of the relationships between all criteria. The only technical aid employed to this end was the recalculation of the weighting values as they were entered and their display in percentage terms (to fit the intuitive description of the process as the allocation of 100 'importance points'). More elaborate multi-criteria techniques (such as various 'swing weighting', 'analytic hierarchy' or 'electre' methods) seek more systematically to build up an overall weighting scheme on the basis of sequential pairwise trade-offs between criteria. However this is achieved only at the expense of significant increases in complexity, with the intervention into the deliberation of a variety of deterministic algorithms and without definitively avoiding potential inconsistencies. The rationale for the straightforward approach adopted in this study rests on the qualified role of the weightings (compared with framing assumptions) and the importance of iterative and reflexive deliberation over ranking.

Finally, the treatment of technical uncertainties in this exercise is by means of deterministic sensitivity analysis (by reference to 'pessimistic' and 'optimistic' scores) rather than the stochastic models employed in *utility function* approaches. The information requirements of such approaches are potentially enormous and always subject to queries over the applicability of the chosen statistically or theoretically-derived probabilities, with results being highly sensitive to a multitude of determining assumptions. In any case, probabilistic methods do not (even in their own terms) offer a valid means to characterise the conditions of strict uncertainty and ignorance which dominate over the formally-defined condition of 'risk' in the case of many aspects of the performance of GM crops.

Beyond these brief remarks relating to key contrasts between the present pilot exercise and the approaches adopted in some other multi-criteria studies, a few comments may be made with regard to certain broad criticisms that are sometimes made of multi-criteria techniques in general.

In commenting positively on the general potential of multi-criteria appraisal, for instance, one recent review for the UK Department of Environment identifies as concerns: the lack of a well-defined procedure for criteria choice; the potential for gaps and overlaps between criteria; the tendency to mix ordinal and cardinal scoring scales and the intrinsic subjectivity of weighting assumptions. Each of these may be taken in turn.

The issue of criteria independence has already been discussed above and in Section 5.5. It will be clear from the discussion in the body of this report that some of the other points might better be seen as advantages rather than as shortcomings in the case of MCM. The lack of constraints on the type of criteria that can be included, the openness to different weighting schemes and the ability to combine quantitative and qualitative factors are all examples in this regard. Indeed, one recent survey of the application of cost-benefit and risk analysis in the energy sector reveals that inconsistent choices of criteria, the existence of gaps and overlaps between criteria and tacit differences of weighting on different factors are at least as much a feature of these approaches.[59] It is only where such issues remain concealed in the presentation of the ostensibly definitive results of risk and cost-benefit analysis that they are truly problematic. Where they are carefully deliberated and openly-declared reflections of particular perspectives in appraisal, then differences of criteria definition and weighting are entirely legitimate. Moreover, it is the flexibility displayed by the explicitly holistic weighting procedures employed in MCM which allows questions such as double counting to be taken into account in the articulation of criteria.

It is for these reasons that the simple linear additive weighting procedure adopted in the present pilot exercise might be argued, on balance, likely to avoid more problems by minimising complexity in analysis than it might solve through any additional formal sophistication.

59 Stirling, 1997.

ANNEX 4

A Technical note on the Treatment of Diversity

The explicit introduction of diversity as a system level consideration using an established index of diversity is a novel feature of this pilot study, justified on the grounds discussed in Section 3.6 (and elaborated in Section 4.8). The potential benefits of diversification across better-performing options as a means to hedge against ignorance and accommodate plural judgements is an issue which is explored in more detail by one of the authors elsewhere.[60] The key idea behind the diversity optimisation technique is that the concept of option diversity is more readily addressed than are the various intractable analytical problems to which this is a response (eg: strict uncertainty, ignorance and value pluralism). Essentially, what is involved is the conditional optimisation *under each perspective* of the degree of diversity in the mix as a whole, subject to a trade-off between a weighting placed on diversity and those assigned to all the various criteria employed in the appraisal of the individual options.

For the purpose of identifying an appropriate measure, the concept of diversity is defined formally as a combination of three subordinate properties: *variety, balance and disparity.* "Variety" reflects the simple number of options in a portfolio. All else being equal, the greater the number of options, the more diverse the portfolio. "Balance" represents the relative importance of the different options in the portfolio. All else being equal, the more balanced the portfolio, the greater the diversity. Finally, the notion of "disparity" addresses the degree to which the different options are qualitatively different from each other. As with any analysis, this is covered in defining the different options themselves— disaggregating them according to their disparity under a range of criteria. Essentially, this is one objective in making the definition of options such an explicit feature in MCM. In all, then, we have in variety, balance and disparity three necessary but individually insufficient conditions for diversity.

Assuming that the property of disparity is addressed in the disaggregation of options (a point taken up below), it is a surprisingly straightforward task to measure the remaining two numerical properties of diversity: variety and balance. A simple algorithm has been developed from first mathematical principles in fields such as statistical mechanics[61] and information theory[62] precisely in order to capture the properties here termed variety and balance.[63] It has been applied as a measure of concentration in economics,[64] of biological diversity in ecology[65] and, most recently, as the measure of portfolio diversity in energy options adopted by the UK DTI.[66] In these latter fields, it is known as the "Shannon-Wiener diversity index". In mathematical notation it may be stated simply as:

$$\boldsymbol{H} = \Sigma_i \, p_i \ln p_i \qquad [3]$$

Where $\boldsymbol{H}$ is the value taken by the diversity index for a mix of options taken as a whole, p_i is the proportional reliance on the i^{th} option and ln is the natural logarithm. Because the logarithms of fractions are always negative, $\boldsymbol{H}$ is always positive. The higher the value of $\boldsymbol{H}$, the greater the diversity.

The business of balancing option performance and portfolio diversity using an index such as this might be termed "*diversity optimisation*". For any given set of options, under any given set of circumstances, there will exist a hypothetical "*diversity optimal portfolio*" with respect to the performance appraisals and ignorance aversion of each participant. In the conventional terms of utility maximisation, this will (in the present case) be the mix of agricultural options for the production of oilseed rape in the UK for which the sum of the utility of the multicriteria

60 Stirling, 1994, 1997, 1999.

61 Betts and Turner, 1992.

62 Shannon and Weaver, 1949.

63 Subject to the specific conditions that the index must take its greatest value for a portfolio of any given variety when the option contributions are perfectly balanced. Second, it must vary monotonically with variety and balance. Third, it should take a minimum value of zero when variety is equal to unity. Fourth, the diversity of a portfolio must remain unaffected if further non-contributing options are taken into account. Fifth, where options are disaggregated according to several independent approaches to classification, the diversity of a portfolio in which options are disaggregated under all classificatory systems must be equal to the sum of the diversities of the same portfolio disaggregated under each individual classificatory system (Laxton, 1978; Betts and Turner 1992).

64 Finkels and Friedman, 1967.

65 Pielou, 1975.

66 DTI, 1998.

performance of the different options and the utility of the diversity of the portfolio as a whole takes some maximum value under a particular perspective. This may be expressed as follows:

$$\max(\boldsymbol{U}); \quad U = \Sigma_i r_i p_i + \delta H \qquad [4]$$

Where $\max(U)$ is the maximum value taken by the total utility U of a portfolio of options, r_i is the utility of the multicriteria performance of option i under a set of weighted appraisal criteria taken from the MCM procedure in equation [1] and p_i is (as in equation [3]), the proportional importance in the mix of option i. The second term is simply the value taken by the Shannon-Wiener index (H) for that portfolio, multiplied by a coefficient (δ) expressing a weighting to reflect the marginal utility of diversity in terms commensurate with the measure of option performance employed in setting r_i. It follows naturally from equation [2][67] that the contribution by each individual option to this optimally diverse portfolio is a simple function of the ratio of the utility of option performance to the marginal utility of diversity for the portfolio as a whole:[68]

$$p_i \propto \exp(r_i/\delta) \qquad [5]$$

Conceived in this way, this 'diversity optimisation' technique offers a potentially useful heuristic means by which to inform decisions over portfolio diversity in MCM. It is relatively straightforward, in that it requires just two basic assumptions additional to those adopted anyway in MCM:

i) that options are defined and disaggregated in such a way as to reflect their disparity;

ii) that it is possible to assign a weighting to diversity in the same way as to other criteria.

A crucial final point may be made in relation to this first assumption concerning disparity. All established analytical approaches to diversity (including scenario and probabilistic approaches), are sensitive to the disaggregation of options. No matter how systematic the treatment, the scheme adopted will always be specific to the context and purpose of analysis and will reflect subjective judgements. Accordingly, there can be no single "objective" taxonomy of disparity against which options may be disaggregated. By adopting the option disaggregation generated by the MCM analysis itself, the present exercise seeks simply to model diversity in the broadest of terms.

The purpose here is the demonstration of the potential merit of this approach as a heuristic and the provisional experimentation with the reactions of participants to the introduction of the diversity issue into appraisal. A more elaborate analysis of the potential role of diversity might be based on the systematic characterisation by participants of their own perspective on option disparities in a fashion analogous to the characterisation of performance in MCM. By addressing the issue of disparity (rather than just the variety and balance components of diversity), such an approach would offer greater completeness than the present exercise. In this regard, one of the present authors has also developed a novel index of 'multicriteria diversity' which, being slightly more elaborate and, at present, not well tested (like the Shannon-Wiener function) has *not* been used in this pilot study. This method is discussed in detail elsewhere.[69] If the property of diversity is judged worth pursuing based on the positive role in the present exercise, then such a multi-criteria diversity index might be applicable in a more elaborate study which builds on the present findings.

67 By means of the method of indeterminate multipliers, subject to the constraint that $\Sigma_i\ p_i = 1$.

68 In fact, it is given by the expression: $p_i = \exp(r_i/\delta) / \Sigma_i \exp(r_i/\delta)$

69 Stirling, 1998b.

References

Andersson, 1988 — R. Andersson, Appraisal of Energy R&D Projects from the Swedish Government's Perspective, Project Appraisal, V3, N1, 3/91, Beech Tree Publishing, Guildford

Arrow and Raynaud, 1986 — K. Arrow, H. Raynaud, Social Choice and Multicriterion Decision-Making, MIT Press, Cambridge MASS

Bana e Costa and Vincke, 1990 — C. Bana e Costa, P. Vincke, Multiple Criteria Decision Aid: an overview, in Bana e Costa, 1990

Bana e Costa, 1990 — C. Bana e Costa, Readings in Multiple Criteria Decision Aid, Springer, Berlin

Barling et al 1999 — D. Barling, The social aspects of food biotechnology: a European view. Environmental Toxicology and Pharmacology. submitted for publication

Beck, 1992 — *Risk Society*, Sage, London

Bell et al, 1977 — D. Bell, R. Keeney, H. Raiffa, Conflicting Objectives in Decisions, John Wiley and Sons, Chichester

Betts and Turner 1992 — D. Betts, R. Turner, *Introductory Statistical Mechanics*, Addison Wesley, Reading

Bezembinder, 1989 — T. Bezembinder, Social Choice Theory and Practice, in Vlek and Cvetkovitch, 1989

Bogetoft and Pruzan, 1991 — P. Bogetoft, P. Pruzan, Planning with Multiple Criteria: investigation, communication, choice, North Holland, Amsterdam

Bohmann, 1996 — J. Bohmann, *Public Deliberation: pluralism, complexity and democracy*, MIT Press, Cambridge

Bonner, 1986 — J. Bonner, *Politics, Economics and Welfare: an elementary introduction to social choice*, Harvester Press, Brighton

Borcherding et al, 1990 — K. Borcherding, O. Larichev, D. Messick, Contemporary Issues in Decision Making, North Holland, Netherlands

Cetron, 1973 — M. Cetron, The Trimatrix: an integratio technique for technology assessment, in Cetron and Bartocha, 1973

Cetron and Bartocha, 1973 — M. Cetron, C. Bartocha, "TA in a Dynamic Environment, Gordon and Breach, London

Chankong and Haimes, 1983 — V. Chankong, Y. Haimes, Multi-objective Decision Making: theory and methodology, North Holland Series in Syste Science and Engineering, Volume 8, North Holland, New York

Chapman, 1981 — M. Chapman, Decision Analysis, Civil Service College Handbook N21, HMSO, London

Citizen's Foresight, 1999 — T. Wakeford, *'Citizen's Foresight: the future of food and agriculture'*, London Centre for Governance, Innovation and Science, University of East London

Clark et al, 1998 — J. Clark, J. Burgess, N. Dando, et al, The Application of Economic Appraisal to Local Environment Agency Plans, Environment Agency, 1998

Clemen, 1991 — R. Clemen, Making Hard Decisions: an introduction to decision analysis, North Holland Series in System Science and Engineering, Volume 8, PWS Kent Publishing, Boston

Climaco et al, 1988 — J. Climaco, A. Gomes Martins, A. Traca de Alemeida, On the Use of Multicriteria Optimization for Energy Planning, International Multicriteria Conference, Paris

Collingridge, 1982 — D. Collingridge, Critical Decision Making: a new theory of social choice, Frances Pinter, London

Covello, et al, 1985 — V. Covello, J. Mumpower, P. Stallen, V. Uppuluri, Environmental Impact Assessment, Technology Assessment and Risk Analysis: contributions from the psychological and decision sciences, Springer / NATO, Berlin

Cropper and Oates, 1991 — M. Cropper, W. Oates, *Environmental Economics: a survey*, Discussion Paper QE90-12-REV, Resources for the Future, Washington

de Bruyn, 1990 — E. de Bruyn, Clinical Decisionmaking in a Multidisciplinary team, in Borcherding et al,

DETR — DETR

DETR, 1998 — D. Pearce, T. Hett, Review of Technical Guidance on Environmental Appraisal, report by EFTEC for Department f Environment, Transport and the Regions, 1998

DTI, 1998 — The Annual Energy Review, HMSO, 1998

Durant and Joss, 1995 — Durant, J., Joss, S., (eds) (1995) '*Public Participation in Science: the role of consensus conferences in Europe*', Science Museum, London

Edwards and Newman, 1982 — W. Edwards, J.. Newman, Multiattribute Evaluation, quantitative applications in the social sciences, Sage, London

EPA,1997 — Presidential / Congressional Commission on Risk Assessment and Risk Management, Framework for Environmental Health Risk Management, Final Report 1997

EPCAG, 1997 — Biotechnology and the European Public Concerted Action Group (1997) Europe ambivalent on biotechnology. Nature 387: 845-847

Field et al, 1994 — F. Field, J. Isaacs, J. Clark, Life Cycle Analysis of Automobiles: a critical review of methodologies, *Journal of Management*

Finkelstein and Friedman, 1967 — M. Finkelstein, R. Friedman, The Application of an Entropy Theory of Concentration to the Clayton Act, *Yale Law Journal*, 76

Fiorino, 1989 — D. Fiorino, 'Environmental Risk and Democratic Process: a critical review', *Columbia Journal of Environmental Law*, 15

Fischoff et al, 1980 — B. Fischoff, P. Slovic, S. Lichtenstein, Labile Values a Challenge for Risk Assessment, Academic Press, London

Friend and Jessop, 1977 — J. Friend, W. Jessop, Local Government and Strategic Choice: an operational research approach to the processes of public planning, Pergamon, London

Giddens A. (1990) — '*The Consequences of Modernity*', Polity, Cambridge

Gonzalez and Tversky, 1990 — R. Gonzalez, A. Tversky, The Impact of Other's Choice on Investment Decisions, in Borcherding et al, 'Contemporary Issues in decision Making', North Holland

Grove-White et al, 1997 — R. Grove-White, P. Macnaghton, S. Mayer, B. Wynne, Uncertain World. Genetically modified organisms, food and public attitudes in Britain. Centre for the Study of Environmental Change, Lancaster University: Lancaster. 64pp.

Hamstra, 1995 — A. Hamstra, *Consumer acceptance model for food biotechnology.* Final Report. SWOKA, Institute for Consumer Research, The Netherlands

Hax and Wiig, 1977 — A. Hax, K. Wiig, The Use of Decision Analysis in Capital Investment Decisions, in Bell et

Hobbs and Horn, 1998 — B. Hobbs, G. Horn, Building Public Confidence in Energy Planning: a multimethod MCDM approach to demand-side planning at BC gas, Energy Policy, V25, N3, pp357-375

Hope and Owens, 1986 — C. Hope, S. Owens, Frameworks for Studying Energy and the Environment, Environment and Planning A: V18

Hope et al, 1988 — C. Hope, R. Hughes, M. Jones, Examining Energy Options Using a SImple Computer Model, Caambridge Univ, Management Stduies Research Paper 14/88, Cambridge University

ILGRA, 1997 — ILGRA, Report on DOE Methodology for Setting Safety Standards, ILGRA/AUG97. 8/97, HM Treasury

Janssen, 1994 — Janssen, Multiobjective Decision Support for Environmental Management, Kluwer, Dordrecht

Japp, 1993 — S. Japp, The Environmental Acceptability of Emerging Technologies, at "Technology and Innovation, 22-24/9/93, Johannesburg, University of Cape Town

Jones et al, 1988 — M. Jones, C. Hope, R. Hughes, A Multi-Attribute Model for the Study of UK Energy Policy, Caambridge Univ, Management Stduies Research Paper 13/88, Cambridge University

Keeney and Nair, 1977 — R. Keeney, K. Nair, Selecting Nuclear Power Plant Sites in the Pacific Northwest Using Decision Analysis, in Bell,Keeney,Raiffa,(ed),Conflicting Objs. in Decs., John Wiley and Sons, Chichester

Keeney et al, 1976 — R. Keeney, H. Raiffa, R. Meyer, Decisions with Multiple Objectives: Preferences and Value Trade-offs, John Wiley, New York

Keeney et al, 1987 — R. Keeney, O. Renn, D. von Winterfeldt, Structuring West Germany's Energy Objectives, Energy Policy, 8/87, Butterworth Heinemann, London

Keeney, 1980 — R. Keeney, Siting Energy Facilities, Academic Press, New York, 1980

Kelly, 1978 — J. Kelly, *Arrow Impossibility Theorems*, Academic Press, New York

Kirkwood, 1982 — C. Kirkwood, A Case History of Nuclear Power Plant Site Selection, J. Operational Research Society, V33, 1982

Kreczko et al, 1987 — A. Kreczko, N. Evans, C. Hope, A Decision Analysis of the Commercial Demonstration Fast Reactor, Energy Policy, 8/87, Butterworth, London, 1987

Lash, Szerszynski and Wynne, 1996 — S. Lash, B. Szerszynski, B. Wynne, *Risk, Environment and Modernity: towards a new ecology*, Sage, London

Laxton, 1978 — R. Laxton, The Measure of Diversity, *Journal of Theoretical Biology*, 70

Lee, 1989 — N. Lee, *Environmental Impact Assessment: a Training Guide*, Occasional Paper 18, Environmental Impact Assessment Centre, University of Manchester

Lubbers, 1989 — F. Lubbers, Planning and Procedural Aspects of the Windfarm Project of the Dutch Elrectricity Generating Board, in Vlek and Cvetkovitch, Social Decision Methodology for Technological Projects, Kluwer, Dordrecht

Luhmann, 1991 — N. Luhmann, *Risk: A Sociological Theory*, de Gruyter, Berlin

MacKay, 1980 — A. MacKay, *Arrow s Theorem: the paradox of social choice - a case study in the philosophy of economics*, Yale University Press, New Haven

Mayer et al, 1996 — S. Mayer, J. Hill, R. Grove-White, B. Wynne, Uncertainty, Precaution and Decision Making: the release of genetically modified organisms into the environment. Global Environmental Change Programme Briefings Number 8. Global Environmental Change Programme: University of Sussex.

McCosh, 1990 — A. McCosh, Supporting Group Decision in Bond Trading, in Borcherding et al, 'Contemporary Issues in decision Making', North Holland, Netherlands

Merkhofer and Keeney, 1987 — M. Merkhofer, R. Keeney, A Multiattribute Utility Analysis of Alternative Sites for the Disposal of Nuclear Waste, Risk Analysis, Vol.7, No.2

Nijkamp et al, 1990 — P. Nijkamp, P. Rietveld, H. Voogd, Multicriteria Evaluation in Physical Planning, North Holland, Amsterdam

NIREX, 1995 — L. Phillipe, Proof of Evidence to the NIREX Public Planning Inquiry, NIREX, 1995

NRC, 1996 — National Research Council, *Understanding Risk: Informing Decisions in a Democratic Society.* National Academy Press: Washington DC.

OECD, 1989 — Organisation for Economic Cooperation and Development, *Environmental Policy Benefits: Monetary Evaluation*, OECD, Paris

ORNL, 1989 — Oak Ridge National Laboratory, William Fulkerson, Charles W. Hagan, et al, Energy Technology R&D: What Could Make a Difference? A Study by the Staff of the Oak Ridge National Laboratory, Volume II: Part 3: X-cutting Sci & Tech, ORNL-6541/V2/P3, Oak Ridge

Pearce and Nash, 1981 — D. Pearce, C. Nash, *The Social Appraisal of Projects: a text in cost-benefit analysis*, MacMillan, London

Pearce and Turner, 1990 — D. Pearce, R. K. Turner, *Economics of Natural Resources and the Environment*, Harvester Wheatsheaf, New York

Phillips, 1989 — L. Phillips, Requisite Decision Modelling for Technological Projects, in Vlek and Cvetkovitch

Pielou, 1977 — E. Pielou, *Mathematical Ecology*, Wiley, New York

Pinkus and Dixson, 1981 — C. Pinkus, A. Dixson, Solving Local Government Problems: practical applications of operations research in cities and regions, George Allen & Unwin, London

RCEP, 1998 — Royal Commission on Environmental Pollution, The Setting of Environmental Standards , HMSO, 1998

Renn, Webler and Wiedeman, 1995 — O. Renn, T. Webler, P. Wiedemann, *Fairness and Competence in Citizen Participation: evaluating models for environmental discourse*, Kluwer, Dordrecht

Renn et al, 1996 — O. Renn, T. Webler, H. Kastenholz, Procedural and Susbtantive Fairness in Landfill Siting: a Swiss case study, Risk, heaqlth, safety and environment, V7, N2, Spring 96 pp95-8, Franklin Pierce Law Centre, Concord

Rip, e al, 1996 — A. Rip, T. Misa, J/. Schot, *Managing Technology in Society*, Pinter, London

Rivett, 1980 — P. Rivett, Model Building for Decision Analysis, Wiley, Chichester

Royal Society (1992) — '*Risk: analysis, perception and management*', Royal Society, London

Royal Society, 1998 — Royal Society, Statement on Genetically Modified Plants for Food Use, 1998

Saaty, 1988 — T. Saaty, Multi-criteria Decision Making: the analytic hierarchy process

Salo, 1995 — A. Salo, Interactive Decision-Aiding for Group Decision Support, Europea Journal of Operational Research, V84, 134-149

SCIMAC, 1998 — Supply Chain Initiative on Modified Agricultural Crops (SCIMAC). Discussion document on Guidelines for growing newly developed herbicide tolerant crops. June

Sclove, 1995 — R. Sclove, *Democracy and Technology*, Guildford Press, New York

Shannon and Weaver, 1962 — C. Shannon, W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana

Sharif and Sundararajan, 1983 — M. Sharif, V. Sundararajan, A Quantitative Model for the Evaluation of Technological Alternatives, Tech. Forecasting and Social Change, V24, 83:15

Smith, 1992 — P. Smith, Testing the Sensitivity of Linear Additive Weighting to Variations in Criterion Weight, Project Appraisal, V7, N2, p107

Solomon and Cameron, 1985 — B. Solomon, D. Cameron, Nuclear Waste Repository Siting: an alternative approach, Energy Policy, 10/85

Starr and Zeleny, 1977 — M. Starr, M. Zeleny, Multiple Criteria Decision-Making, North Holland, Amsterdam

Stewart, 1992 — T. Stewart, A Critical Survey of the Status of Multiple Criteria Decision Making Theory and Practice, International Journal of Management Science, V20, N5-6, 569-586

Stewart, 1996 — T. Stewart, Robustness of Additive Value Function Metods in MCDM, Journal of Multi-criteria Decision Analysis, V5, 301-309

Stirling, 1994 — 'Diversity and Ignorance in Electricity Supply Investment: addressing the solution rather than the problem', *Energy Policy*, **22, 3**

Stirling, 1996 — A. Stirling, On the NIREX MADA, in S. Haszeldine, Radioactive Waste Disposal at Sellafield, University of Glasgow Press

Stirling, 1997 — A. Stirling, Multi-criteria Mapping: mitigating the problems of environmental valuation, in Foster

Foster, 1997 — J. Foster (ed), 'Valuing Nature', Routledge, London

Stirling, 1998 — A. Stirling, Risk at a Turning Point?, *Journal of Risk Research* **1**, 2, 97-110

Stirling, 1998b — A. Stirling, The Economics and Analysis of Diversity, SPRU Electronic Working Paper 28, University of Sussex, December 1998

Stirling, 1999 — A. Stirling, *Science and Precaution in the Management of Technological Risk*, report to the EC Forward Studies Unit, Brussels, May 1999

Suter, 1990 — G. Suter, *Uncertainty in Environmental Risk Assessment*, in Furstenberg

UK National Consensus Conference on Plant Biotechnology, (1994) — *'Final Report'*, London, Science Museum.

UK National Consensus Conference on Radioactive Waste Management (1999), — *'Final Report'*, UKCEED, Cambridge

van den Berg et al 1995 — N. van den Berg, C. Dutilh, G. Huppes, *Beginning LCA: a guide into environmental life cycle assessment*, CML / Unilever / Novem / RIVM, Netherlands

van der Pligt, 1989 — J. van der Pligt, Nuclear Waste: Public perceptiopn and Siting Policy, in Vlek and Cvetkovitch

Vansnick, 1986 — J. Vansnick, On the Problem of Weights in Multiple-Criteria Decision Making (the noncompensatory approach), Euro. J. Operational Research, V24, 86:288

Vlek and Cvetkovitch, 1989a — C. Vlek, G. Cvetkovitch, Social Decision Methodology for Technological Projects, Kluwer, Dordrecht

Vlek and Cvetkovitch, 1989b — C. Vlek, G. Cvetkovitch, Introduction and Overview: social decision making on technological projects, in Vlek and Cvetkovitch

Von Schomberg, 1998 — R. Von Schomberg, *An appraisal of the working in practice of directive 90/220/EEC on the deliberate release of Genetically Modified Organisms*, STOA, European Parliament, Luxembourg, January 1998.

Voogd, 1983 — H. Voogd, Multicriteria Evaluation for urban and Regional Planning, Pion, London

Wathern, 1988 — P. Wathern, *Environmental Impact Assessment: Theory and Practice*, Unwin Hyman, London

Watson, 1981 — S. Watson, Decision Analysis as a Replacement for Cost-beneft Analysis, Europea Journal of Operational Research, V7, 242-248

Webler, et al, 1995 — T. Webler, H. Kastenholz, O. Renn, 'Public Participation in Impact Assessment: a social learning perspective', *Environmental Impact Assessment Review*, **15**, 443-463

Winterfeldt & Edwards, 1986 — D. von Winterfeldt, W. Edwards, Decision Analysis and Behavioural Research, Cambridge University Press

Wynne and Mayer 1999 — B. Wynne, S. Mayer, The Release of Genetically Modified Organisms to the Environment. In "*EU98*. European Environment Agency, Copenhagen, 1999 (in press)