Available online at www.sciencedirect.com

SCIENCE DIRECT®

Technological Forecasting & Social Change 73 (2006) 761–777

Technological Forecasting and Social Change

Evaluation of national foresight activities: Assessing rationale, process and impact

Luke Georghiou *, Michael Keenan

PREST, Manchester Business School, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom

Received 2 May 2005; received in revised form 22 August 2005; accepted 24 August 2005

Abstract

The paper addresses the question of what constitutes an appropriate evaluation strategy for national foresight activities in different situations. The variety of rationales for foresight is explored, ranging from a desire to set priorities through to participation-oriented goals and building new networks around common visions and strategies. A generational model of foresight is used to show the evolution of key evaluation issues. The generic motivations for evaluation of accountability, justification and learning are discussed in the context of foresight. Evaluation grounded in the concept of behavioural additionality and the systems failure rationale is shown to be more suited as a rationale for foresight as public policy.

Assessing the effects of foresight requires an understanding that it is only one of several influences on public policy. To be effective it needs to be tuned into the strategic behaviour and cycles of policy and economic actors. Cases are presented of evaluation of foresight programmes in the United Kingdom, Germany and Hungary. It is concluded that there is no "one-size-fits-all" evaluation approach and that the method selected is conditioned by motivation, timing and the level of aggregation. Foresight cannot be fully evaluated independently from its context. Foresight is being strengthened by the emergence of rigorous and systematic knowledge to assist learning and improvement.

* Corresponding author.
E-mail addresses: Luke.Georghiou@manchester.ac.uk (L. Georghiou), Michael.Keenan@manchester.ac.uk (M. Keenan).

0040-1625/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2005.08.003

1. Introduction

Despite the spread of foresight experience across Europe and beyond, there has not so far been a systematic attempt to understand its effects in aggregate. In particular, foresight has not been evaluated as an instrument of science and innovation policy. Thus far, evaluation has been confined to fairly small-scale panel or survey-based efforts to assess the impact of specific exercises. Comparability across these exercises has neither been sought nor achieved. In this paper we seek to review the experience to date and in so doing to address the question of what constitutes an appropriate evaluation strategy for foresight in different situations. We confine the discussion to national foresight exercises because these are already a large and distinct sub-class and also because they are inter-related through transfer of ideas and practices, sometimes aided by the medium of evaluations and sometimes through other mechanisms.

It is worth considering the relative standing of foresight and evaluation before discussing their interaction. In the array of strategic policy approaches available to the policymaker, the two are often grouped together, with the simple distinction being made that evaluation is looking backwards at what has occurred and foresight looking forward at possible futures. A European working group sought to combine these activities, along with technology assessment, in the domain of technology policy into a common concept of distributed strategic intelligence but largely they remain institutionally distinct [1]. Yet, in other ways, they are not so easily separated, as most evaluations include some form of formative perspective on the future and often have to consider the future implications of the measures they seek to assess. In turn, foresight activity generally needs to be informed by a thorough understanding of the past and present.

However, the object of this paper is to turn one instrument upon the other — to examine how we might go about evaluating foresight and what conclusions that might lead us to in terms of treating foresight as a policy instrument. In many ways the time is right for an evaluative approach. Technology foresight at a national level may now be seen as a policy instrument that is approaching maturity in Europe and Japan at least. Since the early 1990s the practice has diffused widely to the point that most industrialised countries and several advanced developing countries have experience of some form of foresight exercise. Many have been through more than one iteration, while others are about to do so. Collective experience has been accumulated as evidenced not only by the outputs of such exercises but also by a number of efforts to document and analyse what has happened. For example, Durand drew twelve key lessons from the French Key Technologies 2005 exercise [2] and Havas also sought to draw generalisable lessons from the experiences of the Hungarian TEP foresight programme with particular emphasis upon the linkages to policymakers [3]. An earlier generation of programmes was documented in a Special Issue of this journal in which Grupp and Linstone in their synthesis commented that it was difficult to arrive at a general conclusion as to who has benefited from the recent foresight activities the most but that their value as a communication technique appeared to be higher for government and industry than for science organizations [4]. While emphasising the emerging role of foresight as an instrument of communication and benchmarking within innovation systems, the information available to the authors at that time was limited as most of the national programmes had not had time either to achieve their full potential impacts, or to have had them properly assessed.

With the benefit of some additional years of hindsight, in the following sections, after some consideration of what we mean by foresight, the problems and issues involved in its evaluation are discussed. Some experiences of efforts to evaluate individual programmes are considered and some emerging lessons for foresight as a policy tool are discussed.

2. Stated rationales for foresight

Evaluation of policy instruments normally has as a fundamental element a consideration of the rationale for intervention—that is asking why government needs to be taking action in an area. In the traditional framework of innovation policy an intervention should be seen to be correcting a market failure such as asymmetric information, high uncertainty or inability to appropriate the benefits [5]. These arguments have some relevance to foresight in that pooling of knowledge reduces uncertainty yet the benefits are to some extent available to non-participants through published outputs and hence a foresight exercise may not come about without public support. A priority-setting or critical technologies exercise could be justified by arguing that left to the market, players would dissipate their efforts over too wide a spread of activity without reaching critical mass. The intervention of a programme allows network externalities to be achieved, as working within a technology is more likely to be successful if others are working on complementary aspects.

However, the more recent theoretical perspective in the study of innovation, that of system failure [6,7], grounded in evolutionary economics, sits more comfortably as a basis for a rationale for foresight as public policy. As long ago as 1997, Metcalfe [8] argued that foresight (in this case the first UK Foresight Programme) reflected a shift in innovation policy towards "matters of co-ordination, creating and supporting the technology support system of particular groups of firms" that was "entirely consistent with the evolutionary perspective". In the system failure view, the "failures" addressed by public policy may be seen to arise in the rigidities and mistakes of innovation agents (firms, public agencies etc) and in the system itself through a lack of linkages and fragmentation between innovation actors [9]. This perspective also stresses the inter-relatedness between innovation and a series of other factors such as the availability of trained people, finance, and marketing and production capability. The networking benefits of foresight are emphasised. Indeed the possibility of innovation flowing solely from the impetus of a scientific discovery is treated as a special if not rare case (though major discontinuities should be expected from time to time). This move away from the linear model of innovation in itself reinforces the current understanding that foresight needs to be an activity closely engaged with wider socioeconomic trends. Typical questions addressed of foresight as an instrument of innovation policy under this rationale would be:

- Whether the support helps to overcome a lock-in failure by introducing a firm to a new or extended technology or market area? In this case the firm, without the benefit of public intervention through foresight, receives signals only from its own market and existing technological networks and misses the major threats and opportunities coming from beyond these horizons;
- Whether the support is building new networks or coordinating systemic innovations such as those requiring establishment of standards, either between firms or between firms and the science base? Again historical factors may have created certain configurations and relationships which could be deficient in supporting the emergence of a new idea or area. Foresight can be used to build the new social structures, especially in the context of the more distributed and open innovation systems referred to in the definition above.

This rationale may be compared with the typical goals of national foresight activities. Here it is important to stress not only what is common to foresight activities but also the ways in which they differ.

In reality, what we term foresight covers multiple activities and purposes sharing a name. In terms of purpose, some common stated goals for foresight are:

- *Exploring future opportunities so as to set priorities for investment in science and innovation activities*. The degree to which priorities can emerge from foresight varies from "critical technologies" exercises where the whole discourse is focussed on a priority list, through more general programmes from which priorities are derived, to targeted foresight where the priorities are in effect set before foresight begins. As we shall discuss below, the real effect of foresight on priorities may be difficult to determine.
- *Reorienting the Science and Innovation System.* This goal is related to priority setting but goes further. In such cases, there may have been a preliminary diagnosis that the science and innovation system does not match the needs of the country. This was a common situation in Central and Eastern Europe in the immediate post-Communist period when, apart from severe resource difficulties, capabilities reflected an industrial system that no longer existed. Foresight has been used as a tool to re-orientate away from fields such as materials research and towards life sciences as well as to explore new institutional structures.
- *Demonstrating the vitality of the Science and Innovation System.* In this context foresight becomes a "shop window" to demonstrate the technological opportunities that are available and to assess the capability of science and industry to fulfil that promise.
- *Bringing new actors into the strategic debate.* A growing tendency is the use of foresight as an instrument to broaden the range of actors engaged in science and innovation policy. One example is the inclusion of social stakeholders or even sections of the general public such as youth.
- *Building new networks and linkages across fields, sectors and markets or around problems.* A different type of reorientation is sought when foresight is explicitly aimed at creating new networks and or clusters which break out of long-standing disciplinary or sectoral ties.

The modalities of foresight may also differ strongly. All of the above goals may be pursued in a setting which could be organisational, local, regional, or supranational as well as the national level addressed here. The timescale of foresight ranges from the immediate future to the far horizon. The range of actors involved, the process and methods used, and even the status of the activity varies considerably. Foresight ranges from methodological experiment through to major politically driven initiatives. From the perspective of evaluation it is critical to establish which goal, or combination of goals is driving a foresight activity. From the rationale and objectives a framework for evaluation can be established and expected impacts identified. In the next section we consider in more detail how to construct an evaluation approach appropriate to a particular foresight activity.

3. Matching the evaluation approach to rationale and objectives

Against the background of variation in foresight practice described above, one of the more flexible definitions available will be used to help structure the discussion. The origin of the definition is itself evidence of the cumulative learning effects involved in the international diffusion of foresight, having been formulated as part of the efforts of one of the most recent (and smallest) countries to engage in this

activity. It captures key elements of the process that are usually neglected in some of the more commonly used formulations:

> *The foresight process involves intense iterative periods of open reflection, networking, consultation and discussion, leading to the* ***joint refining of future visions*** *and the* ***common ownership of strategies****, with the aim of exploiting long term opportunities opened up through the impact of science, technology and innovation on society…It is* ***the discovery of a common space for open thinking on the future*** *and the* ***incubation of strategic approaches****….* [10]

Of particular importance here is the stress placed upon the way in which joint foresight activities are linked to the joint formulation and ownership of strategies. This perspective avoids what it will be argued below is a common trap, the treatment of foresight and its implementation as separate processes without serious attempts to build bridges between or to link the two. Establishing this connection provides an important backdrop to effective evaluation. This is more than the obvious statement that evaluation of foresight should be concerned with its impacts. Rather, the point here is that foresight is part of a broader set of influences in most of the effects it seeks to achieve in terms of public policy or the strategy of firms. Furthermore, once a foresight output has been produced and enters the environment for implementation the question may be asked of how is it different from other types of policy information emanating from, for example, lobbying, evaluation, strategic studies, or from the influence of historic commitments and budgetary analyses? Probably the answer lies in a longer time horizon, and elements of creativity or commitment through participation, but all of these elements can also come from other sources. The implication is that evaluation of foresight must include understanding of the interaction of foresight outputs with the strategic behaviour of policy and economic actors.

There are also some normative issues involved. Foresight is not always tuned to the needs of recipients and hence, to extend the analogy, the signal may be obscured by noise and not picked up. Information needs to be presented in such a way that policy/strategy mechanisms can receive and absorb it. One moderating factor is that of timing. This needs to synchronise with policy and strategic cycles. Keenan [11] has shown that the results of the first UK Technology Foresight Programme were delivered at the wrong time of year in respect of the priority setting mechanisms of the Research Councils (the funding bodies in the strongest position to implement them), and hence that their impact was both delayed and diminished. Furthermore, the level of recommendations needs to match available funding or capacity for reform. However, foresight cannot always work within the status quo and occasionally it is the policy/strategy structure that needs to change in the light of disruptive foresight information.

If we return to the definition of foresight cited above and combine consideration of this with discussing the implementation issue in foresight evaluation, then it can be argued that the common space and joint ownership elements associated with foresight imply that it should not be viewed as being in a linear or sequential relationship with implementation but rather that it should move into the "implementation space". In other words, the conduct of foresight itself moderates the implementation of emerging findings (or at least the conditions for their implementation), and the environment for implementation affects the way in which foresight ought to be conducted—foresight and implementation are interactive activities. The implication is that, on the one hand impact-oriented evaluation must closely consider the foresight process if impacts are to be fully accounted for and explained, while on the other hand also understanding the drivers of the strategic behaviour of the implementing bodies.

In earlier work one of the authors posited a generational model of foresight which distinguished first generation activity concerned with technological forecasting carried out by futures experts, a second generation bringing in industry and the market as the main actors and third generation foresight adding to these a social and user-oriented perspective [12,13]. Implicit in these models were different approaches to evaluation. For first generation foresight the key issues are accuracy of prediction and diffusion of results (to non-experts). In the second generation the take-up of priorities and establishment of networks among the industrial and academic participants become key evaluation issues, while the third generation implies the involvement of stakeholders in evaluation and looks for evidence of the emergence of a foresight culture. A fourth generation was also suggested in which foresight moved away from the national programme model discussed in this paper and towards a distributed role in the science and innovation system whereby multiple organisations conducted foresight specific to their own needs but with a level of coordination. The implications of the fourth generation for evaluation are discussed in the conclusion to this paper though as yet there has not been an evaluation exercise aimed at such a national system of foresight activities. It is important to note that these generations are ideal types and that an individual foresight activity may exhibit elements of two or even three generations. It does however, emphasise that the approach to evaluation is conditioned by the approach to foresight.

To illustrate this point further, the rationale for conducting a foresight exercise may vary, even within the same exercise. Some possible rationales are shown in Table 1, which clusters the goals set out at the beginning of this section to show expected outcomes and their associated evaluation approaches. Clearly, the evaluator who is looking to attribute resource allocation decisions to a foresight exercise will take a different approach to the evaluator interested in the social network impacts of foresight. That said, more often than not, the professional evaluator will seek to cover all of these 'angles' (and more), not least because of their interdependency (for example, resource allocation decisions are often dependent upon successful problematisation and socialisation of particular issues).

The need for conditioning the evaluation approach to foresight may be illustrated by considering the difference in the unit of aggregation at which foresight is evaluated. For example foresight may be presented as a policy, a programme or as practice. Each of these demands a different evaluation approach. In a policy evaluation, issues of rationale for public action predominate and the interaction of

Table 1
Some examples of rationale for foresight and associated evaluation issues

Rationale for foresight	What does it do?	Expected outcomes?	Evaluation focus?
Providing policy advice	Highlights the longer term and extends perspectives	Policy decisions, resource allocations More rational decision making over space and time	Attribution of decisions to foresight exercise Changes in decision-making processes
Building advocacy coalitions	Highlights challenges in an interaction space around which interest groups coalesce	New emerging networks and communities Wide commitment to realization of a shared vision	Nature of networks Actions undertaken by them
Providing social forums	Provides a 'hybrid forum' for strategic reflection, debate and action	Broadened participation, democratic renewal	Numbers and breadth of actors involved Focus and quality of debates Benefits to participants

foresight with other policies becomes a topic of focus. In the more conventional format of programme evaluation the programme objectives become a primary focus, mostly in terms of objectives achievement but also in terms of the appropriateness of the objectives—the link to policy evaluation. Foresight as practice has as its focus the methods and structures used. These may be evaluated both in their own terms and in terms of whether they were fit for purpose. For example, it is a very different question to ask whether a Delphi survey allows an expert consensus to emerge than it is to judge whether it was the right way to consult a particular group of participants. Again, in a real situation we may see combinations of these elements, albeit with different emphases.

4. Motivations, scope and purpose in the evaluation of foresight

Since foresight is a policy instrument consuming time and resources, it is reasonable to expect that it should be subject to evaluation of a comparable rigour to other tools. In a generalised evaluation framework, three basic tests could be applied:

- Accountability—with questions such as whether the activity was efficiently conducted and proper use made of public funds;
- Justification—with questions such as whether the effects of foresight justify its continuation and extension; and
- Learning—asking how foresight can be done better in particular circumstances.

Each of the above can singly or in combination form the basis of a motivation to evaluate a foresight action and each carries with it implications for the conduct of that evaluation. Although it is the most traditional basis for evaluation, accountability is probably the most problematic in this context, if only because foresight is unlikely easily to yield an answer to the type of question that is being asked here. Accountability generally follows the lines of resources and governance but foresight presents difficulties for evaluators in both cases. Some resources for foresight come centrally from government but in a "soft" policy instrument of this type, the most important resource input is the voluntary contribution of the participants. This is also the case for governance since "ownership" of a foresight exercise is not always clear. Indeed, foresight exercises tend to be distributed activities across (and beyond) communities of actors with various needs and objectives. The practical implication is that evaluators cannot expect to impose demanding questions on a volunteer taskforce that is working without a contract and hence with potentially flexible objectives.

A further difficulty for an evaluation originating in accountability considerations is that there will be pressure to apply standard tests that compare it with other publicly funded activities. Again the specific nature of foresight, as already described, makes this difficult. What happens may ultimately depend upon the degree of institutionalisation of foresight. First efforts in a country or region often represent a "heroic age" in which creativity dominates at the expense of good organisation but where mistakes are forgiven on the basis that this is a prototype. By the time a second cycle is entered institutionalisation may have reversed the proportions of creativity and organisation and brought in its wake requirements for evaluation against more conventional criteria.

The motivation of justification is a common one in evaluation and should not be seen as pejorative. The typical situation is one in which the operators of a programme are approaching a key

decision point regarding extension or repetition of an activity that is likely to be novel and possibly not strongly institutionalised. In these circumstances they desire evidence to support their belief that their efforts have been worthwhile. The independence of view afforded by an evaluation meets this need.

In reality, the justification motive is frequently combined with the third major group, that of learning. Evaluations of foresight are almost inevitably *formative* [14], and given the novelty and sensitivity to context of the instrument, lessons are a likely outcome. These tend to relate to process rather than outcome because of the timing issue which is discussed below.

In a standard evaluation approach, it is important to define the scope and purpose of what is being evaluated at an early stage [15]. The variety of forms of foresight has been discussed in the previous section. Another dimension in which foresight has to be delineated is that of location in time. The key question is where does a foresight activity begin and where does it end. In a first national effort, the beginning is usually clear as the process is initiated with a decision to commit resources and often to establish some sort of secretariat. The end is frequently much less clearly delineated. Where the aim is a report or list of priorities, publication and launch marks some kind of termination though dissemination and other implementation activities may well follow. The launch of networking activities is far less likely to offer a clean break, as these are likely to persist for some time after the foresight activity has ended. An arbitrary decision may need to be made on when to demarcate the cut-off point by when foresight outputs cease to provide a distinct or influential voice in policy discourse. Furthermore, this reinforces the point made in the previous section, that to understand the effect of foresight it is necessary to locate it in a broader strategic and policy context. The evaluation will have to explore the period in which foresight emerged and its interaction with other elements of the system.

The timing issue is also linked to the type of question being asked. If a linear or sequential view of foresight is taken, process issues are best pursued while the activity is still under way. However, many outputs and outcomes will not be clearly visible at this time and will need to be investigated ex post. Here the problem becomes one of attributing effects. Over time, phenomena of 'diminishing signal' and 'knowledge creep' frustrate the evaluator looking to disentangle the effects of foresight exercises from other influences on outcomes. The evaluator needs to move away from such reductionist approaches to a more holistic and systemic framework that views foresight as complementing the use of other policy tools.

Finally, if accuracy of the future visions is an issue the ex post delay corresponds to the foresight period. This may be less of a problem with short horizon five-year critical technology exercises, but it requires a remarkable stable system if the issue is to be usefully pursued for foresight on, say, a fifteen-year timescale. Only the Japanese STA/NISTEP forecasts have been properly assessed on this basis. In this case the method adopted was to use expert panels assembled for the current iteration of Delphi to review the extent to which topics had been "realised" on the expected timescale. The first assessment of this kind was reported in the Fifth Delphi Survey published in 1992 [16]. The panels judged that 28% of the topics had been realised and a further 36% "partly realised". Kuwahara has noted that accuracy in the time dimension is not necessarily the major benchmark as use of the results may have changed the direction of research and technology and affected realisation [17]. Other issues may be raised—accuracy may not necessarily coincide with relevance or utility. Partial realisation is also a difficult concept to enumerate. Data of this kind are best seen as an input to evaluation rather than an evaluation in itself.

5. Criteria and additionality

The classic criteria of evaluation are:

a) *Efficiency of implementation*, otherwise known as process evaluation and focused upon managerial and logistical issues. These are not necessarily trivial or only of bureaucratic concern. Process evaluation covers topics such as organisation and management, and would for example ask: Were the 'right' people involved in an exercise? Did expert panels (if used) receive adequate support? Was the exercise adequately linked to decision-making centres? It may also address the question of the appropriateness and efficiency of methods used, for example: Should a Delphi have been used? Were scenario workshops properly facilitated? A well-conducted process evaluation can cast light upon the dynamics of foresight. As noted above it should be conducted in real-time or immediately after an activity is complete to ensure that the findings are not distorted by hindsight or obscured by loss of data.
b) *Impact and Effectiveness,* often at the core of policymakers concerns, these criteria deal with what has been produced by foresight in terms of outputs and outcomes. Probably the most important observation here is that outputs measure only activity and not its significance. Hence, while it may be useful to know numbers participating in meetings or surveys, reports disseminated, meetings held, website hits and so on, none of these measure the effects of these contacts or their contribution to outcomes. Numbers may even be misleading; the number of "new networks" formed disguises variation in their novelty, size, significance and durability. Outcome evaluation is normally made far more difficult by the problem of attribution, discussed more extensively below.
c) *Appropriateness* as a criterion links back to the earlier discussion of policy level evaluation by engaging with questions of rationale. For national foresight activities this includes the issue of state intervention but also raises questions of what the alternatives would have been (including the counterfactual).

Pursuing this issue of counterfactual, a key question in the evaluation of any public policy intervention is that of additionality—the extent to which the activity would have taken place without a public intervention. This in effect examines the rationale discussed earlier. Under this framework, the questions which should be asked about a foresight activity are:

- Would foresight have happened without the policy intervention?
- Is foresight done differently/better because of the policy intervention?
- Are the resulting actions better because of foresight?
- Have persistent changes been achieved (e.g., foresight culture)?

Within the field of evaluation, recent thinking has moved away from treating additionality as a binary stop-go item. In a "systems of innovation" framework, temporary financial interventions are seen as less important than efforts to change the innovation system for the better in a lasting way. If it is accepted, as discussed above, that foresight is correcting an inherent tendency to have excessively short-term horizons and a difficulty in forming new networks around technologically and socially innovative activities, then foresight may be best evaluated ultimately in terms of its ability to change values and behaviour in these directions.

This perspective on additionality—known as behavioural additionality—has implications for how evaluation is done [18,19]. For example, there is an interest in the persistence of effects—have new routines or practices (including networks developed and the capacity to continue doing foresight) been adopted by the participants? The behavioural additionality perspective is now being applied in many OECD countries to examine the effects of R and D grants for industry. However, it also provides an appropriate framework for evaluating foresight and indeed is being used in practice as the underpinning concept for the current evaluation of the third UK foresight programme. To answer questions of this type we once more return to the need to explore the interactions of foresight with the strategies of the participating organisations. This will require both the practical knowledge of how foresight interacts with and penetrates the organisation, and an understanding of the other influences upon strategic decision-making that contribute to a particular outcome.

The evaluation of foresight should also beware of potential traps. Traditionally foresight is seen as a process of building commitment among stakeholders—an important element for example in Martin's "5 C's" [20]. However, from an evaluation perspective this also creates risks when trying to assess the additionality of foresight. One risk is that of the self-fulfilling prophecy when the "owners" of a foresight activity (for example a sponsor ministry) also control the distribution of resources at the implementation phase. There may be a tendency in this situation to cause foresight priorities to have a stronger influence in the implementation environment than may be justified in terms of the rigour and merit of the exercise. At a more methodological level, stakeholding and consensus may be seen to some extent as a trade-off with creativity and insight. It may be somewhat easier to get "buy-in" to a set of views that are already commonly held than for a really novel or disruptive idea.

6. Some evaluation experiences

Table 2 shows some recent experiences of foresight evaluation, and serves mainly to emphasise that a consistent and comparable approach has not emerged. However, it also demonstrates that

Table 2
Evaluation of national foresight activities

Country	Evaluation effort
Austria	Internal assessment of impacts by Science Ministry
Germany	Delphi 98 evaluation questionnaire; FUTUR evaluated during 2002 and again in 2004
Hungary	Panel evaluation 2003/4 addressing process and impact
Japan	Assessment of realisation of results some 15–20 years after identification in STA forecasts. Also foresight evaluated as a part of broader evaluations of its host institute NISTEP.
Malta, Cyprus and Estonia	"Light" expert evaluation of the eForesee project, examining the achievements of an EU-funded project that linked the foresight activities of these 3 small countries
Netherlands (OCV)	Self-evaluation, PhD study, Masters thesis, evaluation by Advisory Council for Science and Technology (AWT)
Sweden	Process (and not the impacts) evaluated continuously by an evaluation committee. New evaluation in 2005
United Kingdom	Sub-critical ad hoc studies; some limited external (and independent) scrutiny, e.g., by Parliament, a PhD study, etc.

almost every country that has undertaken a national foresight exercise has also seen some need for evaluation. Three of the countries mentioned can be discussed further as case studies. The aim is not to report the result of these evaluations but rather to indicate the types of evaluation approach adopted (and sometimes dropped) and to make some observations about the influence exerted by the evaluations. In effect these are brief meta-evaluations [21,22].

6.1. Case 1: UK foresight evaluation experiences

Rationale and objectives have shifted substantially through the three cycles of the UK Foresight Programme and thinking about evaluation has to some extent tracked this shift. The first cycle was launched as a key policy instrument to reorient the country's science base in the direction of wealth creation and quality of life. The initial expectation was that this would be achieved through establishing priorities for science funding but as time passed and the difficulties of implementing this wish became clear, the secondary objective of building industry-science networks around new opportunities came to predominate. Hence, the programme could be seen to be addressing a failure in the innovation system. In the current third cycle these "second generation" aims survive to some extent but the main focus is upon engagement with stakeholders in government to address policy issues which cut across government structures and have a high science content. The foresight culture objective is also stressed.

Table 3 illustrates the main efforts in foresight evaluation in the UK. It illustrates that without a consistent, credible central approach to evaluation, the likely result is a proliferation of activity. Much of the work listed in Table 3 was at a sub-critical level, or else relied very heavily on anecdotal and potentially prejudiced evidence. The table indicates whether the evaluation got past the stage of planning

Table 3
UK evaluations relating to foresight

Year	Activity	Evaluation implemented	Impact
1995	OST/PREST survey of panelists	Yes	•
1995	OSTS draft comprehensive evaluation proposals that remain unfulfilled	No	•
1995–99	OST sponsors PhD studentship at PREST on evaluation of programme [23]	Yes	•
1996	Panels asked to draft performance indicators	No	••
1996–98	Research Councils and Other Government Departments asked to account for foresight implementation	Yes	•••
1997	Royal Academy of Engineering case study and survey work [24]	Yes	•
1997–2000	Academic evaluations at York [25] and Brunel [26] Universities	Yes	•
1997	Parliamentary Office of Science and Technology produces review of foresight and its impacts [27]	Yes	••••
1997	OST consultation about Lessons from First Round published as Second Round Blueprint [28]	Yes	••••
1998	Segal Quince Wickstead contracted to develop impact indicators	No	•
2000	PREST/WiseGuys/SUPRA contracted to develop evaluation framework for 2nd Foresight cycle	No	•
2001	Chief Scientist's Review of 2nd cycle	Yes	••••
2005	Full evaluation of 3rd cycle commissioned	Yes	Ongoing

•=little or no impact, ••=some impact, •••=significant impact, ••••=major impact.

the methodology and also gives the authors' rating of the impact on the programme. It may be seen that the operating Ministry for foresight, the Office of Science and Technology (OST) was the main driver of activity but, despite commissioning a number of methodological studies and some fieldwork, it has taken over a decade for a full independent evaluation to be commissioned. Impact does not necessarily indicate the quality of the evaluation, only whether its conclusions were implemented. Other interested bodies such as the Parliamentary Office of Science and Technology (POST) were able to put forward more critical and insightful views but lacked the resources to follow up in terms of extensive collection of evidence. The POST report was, nevertheless, unusual in having an impact, despite not being "owned" by OST, partly because of the high status of its Parliamentary home and partly because it was a well-researched and balanced report.

A parallel theme was growing government enthusiasm for performance indicators in all aspects of public sector activity. Notwithstanding the comments made above about the limitations of output indicators, the constant pressure was to capture the effects of foresight in terms of key indicators. During the Second Cycle the authors were asked to develop an indicator-driven evaluation framework. A particular difficulty, as noted earlier, with this accountability-style of approach is that foresight depends heavily upon the unpaid involvement of panelists and other contributors who do not take kindly to being monitored. A "softer" evaluation approach was therefore suggested, which relied upon participants to collect and analyse a significant part of the data, while other items would be compiled centrally. The organising principle was to separate process from impact and in the latter case to identify the five main stakeholder groups: the science base, industry and commerce, the voluntary sector, government, and education, training and public understanding of science. From the framework a set of key indicators was derived that attempted as far as possible to cover the full range of outcomes and outputs. Some of these indicators are outlined in Table 4. In this case impact was curtailed by the early termination of that phase of the programme.

Table 4
Suggested core indicators for the second UK foresight programme

Item measured	Indicator
1. Increased level of awareness of foresight and foresight culture in industry	SME survey questions
2. Commitment of participants	Consistency of attendance at panel and task force meetings
3. Consultation exercise	Number and quality of responses received to consultation documents
4. Cross-panel communication	i) Documented contacts and joint activities ii) Cross-references in reports to issues from other panels
5. Influence on government departments spend	Additional resources committed to foresight activities
6. Influence on government departments coordination	Frequency of foresight on agenda of ministerial science group
7. Influence on science base spend	Proportion of new programmes and initiatives which are clearly aligned with foresight priorities
8. Influence on the formation of new industry–science networks	Persistence of groups founded by panels or task forces
9. Contribution to quality of life goals	i) Engagement of voluntary sector in foresight activities ii) Take-up of recommendations by regulatory or standards-setting bodies in areas such as environment protection, health and safety etc
10. Regional engagement	i) References to foresight in regional innovation strategies ii) Number and extent of regional foresight groups

The stress on accountability may also be seen in the review of the UK's Second Cycle, which superseded the performance indicators framework. The context was a growing realisation that the programme was not on a trajectory likely to offer insights or impacts comparable with its first cycle predecessor. The Government Chief Scientific Adviser was the senior official ultimately responsible for the programme. Recognising the situation, he instituted a high-level review. This was an internally conducted evaluation based upon soliciting views from stakeholders but without any attempt to codify a systematic approach or to present detailed evidence. However, the conclusions were powerful and resulted in a major change of direction that realigned foresight with client interests through a much more focussed approach in which two topics are launched each year such that four run concurrently.

6.2. Case 2: evaluation of German FUTUR initiative

The German FUTUR initiative had a much more specific objective than its UK counterpart in that it sought to introduce fresh ideas into the research funding priorities of one ministry, BMBF, by adding to the traditional mechanisms of agenda setting and prioritisation a third generation approach involving broader stakeholder groups. The recent evaluation of FUTUR [29] was commissioned by the responsible ministry, BMBF and was largely a process evaluation, focusing upon:

- the objectives of FUTUR, which are assumed to summarise the central;
- assumptions upon which the exercise is based;
- the different instruments and methods with regard to their effectiveness;
- efficiency and interplay; and
- the process in general.

The evaluation approach was developed by ISI-Fraunhofer and involved formulating the underlying assumptions and hypotheses that underpin the ideals and conduct of FUTUR. These hypotheses were then 'tested' through their operationalisation into questions that could be detailed in surveys and interview protocols. Following a survey of participants a document was constructed to support an International Panel of Foresight Evaluation Experts. This panel held a one-day hearing with interviews and the Chair (one of the authors) consulted with the Ministry as a user at the most senior level before producing the evaluation report.

In terms of the issues described above, this was an evaluation motivated by justification and learning with the results used both as an input to the decision to continue the initiative into a new phase and in terms of making some adjustments to the process. The limitations of this exercise were too little time and resources available and, in terms of timing, the fact that the exercise was conducted too early to pick up outcomes. However, several process-related recommendations were made and an impetus was gained for the continuation and improvement of the activity. This was an example of foresight being used to bring new actors into the strategic debate and hence one measure of success could be that of participation. On the other hand the quality of that participation and the eventual influence it had on policy outcomes was also an evaluation issue. In this regard a key finding was that the participants felt disconnected from the implementation process and to a lesser extent the programme managers responsible for implementation lacked a sense of ownership of FUTUR, responding more readily to traditional policy inputs and influences. Hence the emphasis upon

understanding the implementation environment and its interaction with foresight processes was highly relevant in this case.

Issues of impact were pursued in more detail in a second evaluation of FUTUR, following broadly the same methodology with a reconstituted panel that met in December 2004 and reported early in 2005.

6.3. Case 3: evaluation of TEP (Hungarian Technology Foresight Programme)

The Hungarian Technology Foresight Programme (TEP) was a holistic exercise with a structure based on sectoral and thematic panels and the use of a Delphi survey, macro-scenarios (an innovative feature in this context) and workshops. The original rationale for the programme was firmly set in the systems of innovation perspective. According to the programme manager the problem faced was a highly fragmented innovation system during a time of fundamental economic and social change [30]. The Programme aimed to bring together business, the science base and government to identify and respond to emerging opportunities in markets and technologies. The evaluation [31] was commissioned by the successor body to the ministry that had sponsored the programme (the Research and Development Division of the Ministry of Education) three years after completion of the main phase of the TEP. Its objectives were:

i) To answer the question as to what extent TEP has achieved its objectives; and
ii) To help orient decision-making on future foresight activities in Hungary.

In so doing it was to address:

- appropriateness of the original objectives;
- whether the original objectives were adequately formulated;
- organisational structure and performance of the management of the exercise;
- the extent to which direct and indirect outputs have been achieved;
- justification of the exercise in terms of value for money;
- barriers to implementation of the recommendations;
- recommended directions for the organisation of future foresight activities in Hungary.

The approach adopted was again a light one, with an international panel designing a questionnaire for stakeholders (two thirds of whom were panelists or members of the overall Programme Steering Group, with the rest being experts or government officials) and hearings during which 22 participants and users were interviewed by the panel.

Again this could be seen as a mixture of learning and justification, though another motivation was to revive interest in foresight after a political hiatus and in the formative element of the evaluation to assist in the design and adoption of a new type of foresight activity. The later timing was very beneficial as many of the effects had taken much longer than expected to materialise and were recent in terms of the evaluation. TEP provides a good example of the importance of policy context, as the initial failure to have a significant impact resulted mainly from a lack of a clear client base—the sponsor organisation OMFB was not at that time well-connected to propagate results across government and the situation was worsened by political change.

In the event, the evaluation found that that the most important effects were mainly in the area of cultural change—in establishing longer-term perspectives and in introducing greater inter-disciplinarity. This was seen as a welcome introduction of longer-term thinking during a period when the country was dominated by a short-term agenda (partly because of economic difficulties but also as a reaction against long-term planning of a previous era). Questionnaire respondents were quite negative about the effects achieved in terms of the original objectives, particularly in influencing the research directions of industry or the public sector. As noted above, the panel found that effects on public policy had been substantial but had materialized through a slow and non-linear process.

7. Conclusions

Examining foresight against the type of framework presented by evaluation raises several issues. If we consider the range of situations in which foresight is applied, ranging from priority setting in technology strategy, through network-building and participative approaches to broader restructuring of the science system it is evident that there is no "one-size-fits-all" evaluation approach. Furthermore, the evaluation approach is affected by a series of other factors, including the motivation for evaluation, its timing in relation to foresight activity and the unit of aggregation at which foresight is approached.

The conceptual basis of evaluation helps to place some order. A systematic understanding of the rationale behind a particular foresight intervention can in turn lead to an evaluation framework. However, the systems rationale also leads us to the realisation that foresight cannot be evaluated independently of its context. This falls out at two levels: the need to understand the relative signal strength of foresight compared with other influences in determining the attribution of impacts, and the interaction of foresight with the strategies of the organisations it seeks to affect. Evaluation has to steer a difficult course between under and over-attribution.

The three national case-studies discussed illustrate the variety of evaluation approaches but also indicate the immaturity of evaluation in this context. Process issues and early impacts are generally well taken care of by what is effectively peer review but much remains to be done in understanding and measuring impacts, particularly in the longer term. As many of the countries discussed here move towards the fourth generation "distributed" model of national foresight, some new evaluation issues appear. On the one hand the political pressures surrounding a national initiative of high prominence are less likely to be present, which in principle should make evaluation easier to carry out. On the other hand, there is an increased complexity as it is likely that policymakers will want to understand the linkages between different future-oriented exercises and their sources of information and expertise. Another trend has been the emergence of several exercises in Europe which aim to systematise the exchange of knowledge about foresight and its effects, either through information platforms or through explicit approaches to benchmarking. However, attaching credibility and validity to such activities places rigorous evaluation as a prerequisite. Foresight can only be strengthened by the emergence of a more systematic and rigorous body of knowledge to assist learning and improvement. Furthermore, as with foresight itself, evaluation can offer a forum for the exchange of knowledge that allows some of the more tacit but essential lessons to diffuse through the community.

References

[1] S. Kuhlmann, et al., Improving distributed intelligence in complex innovation systems. Final report of the Advanced Science and Technology Policy Planning Network (ASTPP), Karlsruhe/Brussels 1999 (ISI/ European Commission).
[2] T. Durand, Twelve lessons from 'Key Technologies 2005': the French technology foresight exercise, J. Forecast. 22 (2003) 161–177.
[3] A. Havas, Evolving foresight in a small transition economy, J. Forecast. 22 (2003) 179–201.
[4] H. Grupp, H.A. Linstone, National technology foresight activities around the globe—resurrection and new paradigms, Technol. Forecast. Soc. Change 60, 85–94 (1999) p. 92.
[5] K. Arrow, Economic welfare and the allocation of resources for invention, in: R. Nelson (Ed), The Rate and Direction of Inventive Activity, (Princeton: Princeton University Press), 1962, 609–625; re published in N. Rosenberg, The Economics of Technological Change (Harmondsworth: Pelican), 1974, 164–181, 1962.
[6] OECD, Technology, Productivity and Job Creation — Best Policy Practice, OECD, Paris, 1998.
[7] R. Lipsey, Technology policies in neo-classical and structuralist–evolutionary models, STI Rev. (22) (1998) 31–73.
[8] J.S. Metcalfe, Science policy and technology policy in a competitive economy, Int. J. Soc. Econ. 24 (789) (1997 September).
[9] K. Smith, Innovation as a systemic phenomenon: rethinking the role of policy, Enterp. Innov. Manag. Stud. 1 (1) (2000) 73–102.
[10] J. Cassingena Harper (Ed.), Vision Document, eFORESEE Malta ICT and Knowledge Futures Pilot, 2003 (January).
[11] M. Keenan, An evaluation of the implementation of the UK Technology Foresight Programme, Doctoral Thesis, PREST, University of Manchester, 2000.
[12] L. Georghiou, Third generation foresight — integrating the socio-economic dimension, Proceedings of the International Conference on Technology Foresight — The Approach to and Potential for New Technology Foresight, NISTEP Research Material, vol. 77, 2001 (March).
[13] L. Georghiou, Foresight concept and practice as a tool for decision-making, Plenary presentation, Technology Foresight Summit, Budapest, Hungary 27–29 March 2003, UNIDO, Vienna 2004 and http://www.unido.org/file-storage/download/?file_id=10631.
[14] M. Scriven, Evaluation Thesaurus, 4th ed., Sage, Beverly Hills, CA, 1991.
[15] M. Gibbons, L. Georghiou, Evaluation of Research — A Selection of Current Practices, OECD, Paris, 1986.
[16] NISTEP, 2nd Policy-oriented Research group, 5th technology forecast survey — future technology in Japan, NISTEP Report No. 25, Tokyo, 1992.
[17] T. Kuwahara, Technology forecasting activities in Japan, Technol. Forecast. Soc. Change 60 (1999) 12.
[18] T.J. Buisseret, H.M. Cameron, L. Georghiou, What difference does it make? Additionality in the public support of R and D in large firms, Int. J. Technol. Manag. 10 10 (46) (1995) 587–600.
[19] L. Georghiou, Impact and additionality of innovation policy, in: P. Boekholt (Ed.), Innovation Policy and Sustainable Development: Can Innovation Incentives Make a Difference, IWT-Observatory, Brussels, 2002, pp. 7–22.
[20] B.R. Martin, R. Johnston, Technology foresight for wiring up the national innovation system—experiences in Britain, Australia, and New Zealand, Technol. Forecast. Soc. Change 60 (1) (1999 January 2) 37–54.
[21] D.L. Stufflebeam, The metaevaluation imperative, Am J. Eval. 22 (2) (2001) 183–209.
[22] M. Scriven, An introduction to meta-evaluation, Educational Product Report No. 2, 1969, pp. 36–38.
[23] M. Keenan, An evaluation of the implementation of the UK Technology Foresight Programme, Doctoral Thesis, PREST, University of Manchester, 2000.
[24] Royal Academy of Engineering, A Preliminary Assessment of the Impact of Foresight, RAE, London, 1998 (June).
[25] A. Webster, Technologies in transition, policies in transition: foresight in the risk society, Technovation 19 (1999) 413–421.
[26] M. Henkel, Academic responses to the UK foresight programme, High. Educ. Manag. 12 (1) (2000) 67–84.
[27] United Kingdom Parliamentary Office of Science and Technology, Science Shaping the Future: Technology Foresight and its Impacts, POST, London, 1997 (June).
[28] Office of Science and Technology, Blueprint for the Next Round of Foresight, available at http://www.foresight.gov.uk/Previous_Rounds/Foresight_1994_1999/Blue_Print_for_the_Next_Round_of_Foresight.html last accessed 19 August 2005.
[29] K. Cuhls, L. Georghiou, Evaluating a participative foresight process: 'FUTUR—the German research dialogue', Res. Eval. 13 (3) (2004 December) 143–153.

[30] A. Havas, Evolving foresight in a small transition economy, J. Forecast. 22 (2003) 179–201.
[31] Evaluation of the Hungarian Foresight Programme, Report of an International Panel, May 2004, Hungarian National Office for Research and Technology, http://www.nkth.gov.hu/main.php?folderID=159&articleID=3826&ctag=articlelist&iid=1 last accessed 19 August 2005.

Luke Georghiou B.Sc., Ph.D., Professor of Science and Technology Policy and Management at PREST, Manchester Business School, University of Manchester, United Kingdom.

Michael Keenan B.Sc., Ph.D., is Lecturer in Foresight at PREST, Manchester Business School, University of Manchester, United Kingdom.