Technological Forecasting & Social Change 76 (2009) 869–879

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

The UK technology foresight programme: An assessment of expert estimates

Felix Brandes *

TNO, Innovation Policy Group, Schoemakerstraat 97, 2628VK, Delft, the Netherlands

ARTICLE INFO

Article history:
Received 6 August 2008
Received in revised form 27 January 2009
Accepted 28 January 2009

Keywords:
Foresight
Forecasting
Delphi
Survey
Expert panels
Evaluation
Optimism
Gaming

ABSTRACT

Since the early 1990s, 'Technology Foresight' exercises with special emphasis on the use of Delphi surveys have played an important role in science and technology (S + T) policy across Europe in an effort to focus resource allocation. Yet, none of the estimates made in the European Delphi surveys have been formally assessed in retrospect, while this process has been incorporated into the Japanese surveys since 1996. Taking the UK Technology Foresight Programme, this research sets out to assess the estimates of three of the fifteen panel Delphi surveys. Whilst on average 2/3 of Delphi statements were predicted to be realised by 2004, it will be shown that only a fraction of these statements had been realised by 2006. Based on the evidence collected from the published panel reports, the 'Hindsight on Foresight' survey conducted by OST in 1995 and interviews with panel members, it will be argued that the overwhelming majority of estimates were overly optimistic. While optimism and strategic gaming of experts is the most convincing explanation for these results, process factors were also explored, including the quality of expert panels used, the Delphi statements and the respondents of the Delphi questionnaire. It is argued that at least the issue of short-range optimism and strategic gaming of experts should be addressed in future Delphi exercises, as decision makers relying on expert advice cannot deal with this issue alone.

1. Introduction

In the early 1990s many Western European countries experienced a "revival" of national 'technological forecasting' or 'foresight' initiatives in an attempt to focus resource allocation in S + T policy. Until then Japan was the country with the longest experience in this field, having conducted quinquennial 'Delphi' surveys since 1971 [1]. Consequently, the Europeans sought advice from the Japanese with the Germans translating the 5th Japanese Delphi survey and re-running it in Germany in 1993. Other countries such as France soon followed down that road [2]. However, not all European countries chose Delphi as their preferred method as especially smaller countries like the Netherlands used other approaches instead [2].

The UK embarked on their Technology Foresight Programme as recommended in the 1993 White Paper 'Realising our Potential', employing 15 expert panels representing the UK's commercial and technological base, supported by a large scale national Delphi survey. The survey was sent to 8384 experts in 1994, of whom 2585 participated, to generate estimates with a time horizon of 2015 or beyond [3]. However, across panels 2/3 of statements were predicted to be realised between 1995 and 2004 [3]. Thus, the main objective of this paper is to assess the 1994 estimates for their current status of realisation to see how realistic the expert estimates were.

This objective might be contested by some readers as they might argue that it is irrelevant whether the estimates have turned out to be correct, as the emphasis of the exercise should be on whether the estimates at the time have enabled good decision making [4]. Furthermore, the most accurate forecast is not necessarily the most useful one. Forecasts are at times most effective if they are self-fulfilling or self-defeating [5]. This is why it is not possible to measure the success of Delphi studies based on the share of realised statements. This is certainly true, and it is not the intention of this paper to judge the success of the UK Delphi survey as

* Tel.: +31 152 695 432; fax: +31 152 695 460.
E-mail address: felixbrandes@gmail.com.

0040-1625/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2009.01.005

many more factors beyond the estimates determine the success of a Delphi. Nevertheless, there are a number of important questions to ask about future estimates:

1) Did the forecasted items come about?
2) If not, have the items still chance of occurring, or did they occur earlier than expected?
3) What items occurred that were not forecasted?

This paper only deals with the first category as to this date very few Delphi surveys conducted in Europe have been evaluated over the last decade. However, this should only be seen as the first step with need for future research addressing the other questions. This is important as recommendations and decision making were at least partially based on the UK Delphi estimates. If estimates turn out to be consistently over-optimistic or incorrect this should be taken into account or the method improved to reduce this bias. An important issue raised by this analysis, although contested, is the question what percentage of items one would expect to be realised i.e. what target level does one expect? As the future is uncertain a 100% realisation rate is unrealistic and undesirable. However, with few studies apart from the Japanese analysed from hindsight this is difficult to say. For this paper the realisation rates of the Japanese Delphis are used for comparison.

2. Use of Delphi in science and technology policy

The Delphi technique as conceived by the RAND Corporation in the 1940s and 1950s has spread widely into policy, industry and academia alike [6]. Because of its wide application and many variations, today most practitioners probably agree that the Delphi method has two irreducible elements: anonymity and feedback between rounds [7]. The method has therefore evolved substantially from its original form, where for example the importance of consensus judgements was emphasised with the 'Policy Delphi' first diverting from this original format [8].

The first (inter)national scale Delphi survey was conducted in the US in the 1960s [9]. The lack of application of results in the political system and oil crisis greatly reduced the perceived need for such future studies, which meant that the US has not conducted a national Delphi since [1]. Instead it was Japan who championed the Delphi method at a national scale in S + T policy, conducting surveys ever since 1971. With the renewed interest in 'technological forecasting' in Europe in the early 1990s, national technology Delphis spread in Europe. While Germany, France and the UK were the first to conduct national Delphi surveys, today the main interest lies in Central and Eastern Europe. Contrary to the originally designed purpose of generating estimates, the main reason for using national Delphis is often argued to be its role in connecting the various actors in the National Innovation System [6]. Nevertheless, most of the national surveys conducted, generate quantitative estimates usually on a range of variables. As this is the main purpose of Delphi surveys – the networking role is a secondary aspect which can be replaced by other methods such as standard surveys – it is useful to analyse the estimates in retrospect as the estimates are designed to support decision-making processes. Recently, it seems that there is some dissatisfaction with the results of Delphi surveys in the countries first adopting it. Germany, France and the UK have abandoned the use of national scale Delphis in their foresight exercises, with even Japan switching from the sole use of Delphi surveys and supporting its latest Delphi using scenarios, bibliometrics and other surveys [10]. Instead, most national foresights use a combination of methods, now, often focused on a specific topic, compared to the broad national Delphi surveys of the 1990s.

3. Brief introduction to the UK Delphi

In 1993 the Office of Science and Technology (OST) White Paper: 'Realising our potential' reaffirmed the need for priority setting to make the UK science base more responsive to industry needs [11]. As a result the first holistic overview of S + T priorities in the UK was attempted, embarking on the UK Technology Foresight Programme. The main goal of the programme was to identify future S + T areas, which could be exploited for wealth creation and improvements in quality of life (ibid). In addition new partnerships between industry and scientists were hoped to be forged [12].

Fifteen panels, representing the different UK industry sectors, were set up to produce a report on future S + T issues in their sectors. To do so panel members received basic training on the methods and concepts that they would be expected to employ during the main foresight stage, which were scenario-building, the use of a Trends and Issues survey, Delphi, and the use of prioritisation criteria [13]. The Delphi survey was used as it was felt that the necessary expertise was not accessible within panels of manageable size [12]. Furthermore, to ensure that the findings would be implemented afterwards it was deemed necessary to engage "with the widest possible cross-section of experts in the UK" [12]. The Delphi survey was therefore also intended as a "communication" tool, rather than merely predicting variables on future generic technologies.

Nevertheless, Georghiou stated that the survey was conducted to inform the panels' views about future developments based on estimates by the relevant expert communities [12]. Therefore each panel developed eighty statements grouped into several sub-sectors, which experts were asked to make estimates about. While the statements were developed based on a separate survey, as well as regional workshops in the case for some panels, an unstructured round conducted with the later response group to generate the statements was not part of the methodology. The original questionnaire was of the following format. Experts were asked to rate their level of expertise on a scale from one to five for each of the statements. One representing a level of expertise defined as "Unfamiliar" and level five as "Expert". Exact definitions of each level were specified on the questionnaire [3]. Respondents were further asked to make estimates in regard to (1) the expected impact on wealth creation and quality of life,

(2) the expected period of occurrence, (3) the degree of collaboration required for development, (4) the UK's current position in the field compared to other countries and (5) the expected constraints on the occurrence in the UK [3].

While the Programme suffered from severe time pressure as Government ministers set tight deadlines, the overall Foresight Programme was perceived as a success with the panel reports well received with little criticism [13]. The main criticism centred on methods employed with particular focus on the Delphi, which suffered most from time pressure. The Health & Life Sciences Panel stated that its Delphi statements were poor[1] as they were developed in a single afternoon and consequently did not influence the final panel report. However, Georghiou and Keenan argue that this criticism of the Delphi should be seen in context, as it was not the quality of the statements but the time pressure that prevented the effective use of Delphi results, especially the second round results [13]. Additionally, the unfamiliarity with the method also won the Delphi few friends amongst the panel members.

While the Delphi was widely criticised at the time for the reasons mentioned above, even leading to questions raised in the House of Commons [11], assessing the Delphi estimates from retrospect should enable a more informed judgement.

4. Methodology

As the survey for this research received no funding it had to be conducted using a most cost effective methodology. Consequently, not all 15 panels could be assessed but instead only a selection. It was attempted to choose 3 to 4 panels representing the different characteristics of the overall panels in regards to short and long term innovation sectors as well as manufacturing and service sectors. The selection of the 'Chemicals', 'Energy' and 'Retail & Distribution' panel was thought to represent these criteria.

To assess the current status of realisation of the 1994 Delphi statements, experts had to be consulted for their judgement. As it was difficult to use small expert groups assessing the status of realisation in practice, it was decided to use a semi-structured (online-)questionnaire instead. Because the original Delphi questionnaires were criticised to be too lengthy, as well as experts feeling uncomfortable to make judgements on such wide ranging issues,[2] the original 80 statements were grouped into 3 to 4 shorter questionnaires. To prevent any bias the original statements were used unaltered only inquiring whether they had to date, i.e. 2006, been 'Realised', 'Partially Realised' or 'Not Realised'.[3] A further 'Don't Know' option was given to prevent experts from being forced to make a judgement. Additionally, an open comment box was provided for each statement for experts to explain their judgements, if necessary, of which extensive use was made.

Generally, it was attempted to include experts from industry, public services and academia. In total 126 survey invitations were sent to experts from the chemicals, energy and retail& distribution sector. Overall, a response rate of 38% was achieved, considered high for an online survey. However, the response rate varied for the different panels with the highest response rate of 46% for the retail & distribution panel, 35% for the energy panel and 33% for the chemicals panel.

After collecting the necessary data for an assessment of the estimates, the results had to be interpreted. While the 'Hindsight on Foresight' survey with its section on the Delphi survey was used as a rich source of information, it was felt that the results should be presented to original panel members for their interpretation. Consequently, it was attempted to interview at least one expert from each of the assessed panels via telephone. While this was possible for the 'Chemicals' and 'Retail & Distribution' Panel, this was not achieved for the 'Energy' Panel. Furthermore, the 1995 panel reports were used as further aid in the interpretation as well as comparing the results to past Japanese experiences. Overall, this methodology allowed exploring process factors that might explain the survey results at a minimum cost. External factors such as barriers to the development of technologies could not be explored with this methodology. Nevertheless, it is felt that the conclusions drawn are valid taking into account the limitations acknowledged.

5. Presentation and analysis of survey results

The overall survey results regarding the status of realisation of Delphi topics in 2006 are displayed in Table 1. Whereas all three assessed panels developed eighty statements, the number of estimates falling into the 1995–2004 time-band and hence the rate of topics expected to be put into practice[4] by 2004, varies considerably between the panels. 'Chemicals' has the lowest 'Expected

[1] According to the organisers of the Delphi this was due to the chaotic organisation and poor chairmanship of the panel.

[2] These comments were voiced in the 'Hindsight on Foresight' survey conducted by OST after the first stage of the UK Technology Foresight Programme in 1995.

[3] The same definition of terms were used as in the Japanese assessments. Namely
Realised – statement has been realised at a past date (in this case 2006).
Partially Realised – an aspect of the statement has been realised at a past date, these are:

- Cases where a single topic forecasts two or more aspects, and while one or more aspects have been realised, there remains at least one aspect that has not been realised.
- Cases where an expression (including adjectives describing performance) in the topic is not quantitatively defined, and its status of realisation is open to interpretation.
- Cases where a part of the requirement described in the topic has been realised.

Not Realised – the statement in question has not been realised up to the present.

[4] Put into practice' is here defined as the realization of the topic statement. The original Delphi survey statements used the words 'elucidation', 'development', 'practical use' and 'widespread use' to delineate different stages in the development cycle. Experts assessed the full, original Delphi statements taking into account these wordings. 'Put into practice' is hence not to be confused with widespread use but refers to the wording used in the original topic statement.

Table 1
Overall presentation of survey results.

Sector	Number of topics	1995–2004 prediction	Expected realisation rate [a] (%)	Realised		Joint realisation rate [b] (%)	Not realised		
Chemicals	80	41	51%	4	5%	28%	51	64%	
Environment and safety	17	13	76%	0	0%	24%	13	76%	
Catalysis	8	5	63%	1	13%	38%	5	63%	
Chemical processes	12	5	42%	0	0%	17%	9	75%	*
Information and design systems	7	4	57%	2	29%	29%	4	57%	*
Chemical products and materials	11	7	64%	1	9%	45%	5	45%	*
Sensors and measurement	7	3	43%	0	0%	43%	4	57%	
Energy and feedstocks	6	0	0%	0	0%	17%	4	67%	*
Education and training	2	0	0%	0	0%	0%	2	100%	
Healthcare and biotechnology	10	4	40%	0	0%	20%	5	50%	*
Energy	80	50	63%	12	15%	34%	44	55%	
Discovery, extraction and supply	26	17	65%	5	19%	54%	10	38%	*
Transmission, transportation and distribution	12	7	58%	0	0%	17%	7	58%	*
Conversion	15	9	60%	4	27%	33%	9	60%	*
End use	27	17	63%	3	11%	22%	18	67%	*
Retail & Distribution	80	60	75%	5	6%	43%	41	51%	
UK social trends	15	12	80%	2	13%	33%	8	53%	*
Product related	6	4	67%	0	0%	17%	5	83%	
Supply Chain	13	9	69%	0	0%	54%	6	46%	
Marketing	7	3	43%	0	0%	57%	3	43%	
Distribution	10	10	100%	0	0%	30%	7	70%	
Consumer	9	6	67%	1	11%	44%	5	56%	
Retailer	15	12	80%	0	0%	47%	7	47%	*
Financial services	5	4	80%	2	40%	60%	0	0%	*

*indicates sub-sectors containing 'No clear Assessment' topic statements.
[a] The 'Expected Realisation Rate' is the percentage of total statements estimated to come about in the time period 1995–2004.
[b] The 'Joint Realisation Rate' is the sum of percentages of estimates assessed as 'Realised' and 'Partially Realised'.

Realisation Rate' of 51%, whereas 'Energy' contains 63% and 'Retail & Distribution' 75% of topics predicted to be realised by 2004. While these are the expected average percentages for the overall sectors, the estimates vary widely between the sub-sectors. For example, the sub-sector of 'Distribution' contains only 1995–2004 estimates, whereas some 'Chemicals' sub-sectors contain no 1995–2004 estimates.

One major reason to assess the topics for their status of realisation at this point in time – after 12 years only – is the large number of topics predicted to come about in the first decade. However, as the column of 'Realised' topics shows, only very few estimates have been realised as predicted. Overall, only 4 (5%) of the 'Chemicals' statements and 5 (6%) of the 'Retail & Distribution' statements were assessed as 'Realised' by the experts. The estimates of the 'Energy' survey fare much better compared to the other two surveys with 12 (15%) topics fully realised to date. However, the current status is nowhere near the 1995 estimates, which consequently have to be classified as overly optimistic. This over-optimism of experts is not surprising looking at past experiences.

As we deal with an uncertain future one would not expect all topics to materialise as predicted in terms of timing and scope. As a result the category of 'Partially Realised' topics was used in the survey, to account for statements which have undergone development in some aspect. The 'Joint Realisation Rate' is the percentage of topics which were assessed as either fully or partially realised to date, which gives a more favourable assessment as it includes estimates which have only partially materialised. While 'Chemicals' has the lowest 'Joint Realisation Rate' of 28%, 'Energy' has a 34% and 'Retail& Distribution' a 43% rate. Surprisingly, the large number of fully realised 'Energy' statements is not reflected in a proportionally higher 'Joint Realisation Rate'. This anomaly will be discussed in more detail in the discussion section.

The column of 'Not Realised' statements is presented as an indicator of how many topics have not been put into practice to date. 'Chemicals' contains the most 'Not Realised' topics with 64%, while 'Energy' contains 55% and 'Retail & Distribution' still contains 51% of 'Not Realised' topics. Consequently more than half of all statements have not undergone major development since 1995 as indicated by partial realisation. The (*) in the final column indicates sub-sectors that contain statements where experts made contradicting assessments which overall had to be classified as unclear. The sum of 'Joint Realisation Rate' and 'Not Realised' rate for these sub-sectors are therefore not adding up to 100%.

5.1. Analysis of 1995–2004 estimates

While Table 1 presented the overall results of the survey, Table 2 looks at the sub-group of 1995–2004 estimates in detail as the above analysis includes topics assessed as fully or partially realised with predicted realisation horizons beyond 2004, and hence distorting the results.

Table 2
Analysis of 1995–2004 estimates.

Sub-sector	Number of topics (assessed)	Realised (%)	Joint realisation rate (%)	Not realised (%)
Chemicals	41	2%	29%	56%
Environment and safety	13	0%	23%	77%
Catalysis	5	20%	60%	40%
Chemical processes	5	0%	20%	60%
Information and design systems	4	0%	0%	75%
Chemical products and materials	7	0%	43%	43%
Sensors and measurement	3	0%	67%	33%
Energy and feedstocks	–	–	–	–
Education and training	–	–	–	–
Healthcare and biotechnology	4	0%	0%	25%
Energy	50	24%	44%	42%
Discovery, extraction and supply	17	29%	65%	24%
Transmission, transportation and distribution	7	0%	0%	71%
Conversion	9	44%	56%	44%
End use	17	18%	35%	47%
Retail & Distribution	60	8%	43%	50%
UK social trends	12	17%	42%	50%
Product related	4	0%	0%	100%
Supply chain	9	0%	56%	44%
Marketing	3	0%	33%	67%
Distribution	10	0%	30%	70%
Consumer	6	17%	67%	33%
Retailer	12	0%	50%	42%
Financial services	4	50%	50%	0%

The data in Table Table 3 supports and even amplifies the previous results. Looking at this sub-group of estimates, only 1 (2%) of the 1995–2004 'Chemicals' estimates has been put into practice, whereas 5 (8%) of 'Retail & Distribution' estimates have come about. As before, the 'Energy' estimates stand out from the rest with 12 (24%) fully realised topics and 44% 'Joint Realisation Rate'. Again, only few of the estimates have been fully put into practice, while a larger share of 1995–2004 estimates are at least partially realised, reflected in the 'Joint Realisation Rates'. The 'Energy' panel is the only one containing less than 50% of 'Not Realised' topics, which acts as further evidence that the 'Energy' estimates are of higher quality than the rest.

Table 3
Realisation rates by predicted time band.

	Number of topics	Realised (%)	Joint realisation rate (%)	Not realised (%)
Chemicals	80	5%	28%	64%
1995–1999	0	–	–	–
2000–2004	41	2%	29%	56%
2005– 2009	25	8%	28%	68%
2010– 2014	3	0%	0%	100%
2015 or beyond	3	0%	33%	67%
Never	5	0%	20%	80%
Bimodal	3	33%	33%	67%
Energy	80	15%	34%	55%
1995–1999	7	29%	29%	43%
2000–2004	43	23%	47%	42%
2005–2009	20	0%	15%	75%
2010–2014	2	0%	0%	100%
2015 or beyond	4	0%	25%	75%
Never	0	–	–	–
Bimodal	4	0%	25%	75%
Retail & Distribution	80	6%	43%	51%
1995–1999	20	10%	55%	35%
2000–2004	40	8%	38%	58%
2005–2009	9	0%	33%	56%
2010–2014	1	0%	0%	100%
2015 or beyond	0	–	–	–
Never	4	0%	0%	100%
Bimodal	6	0%	83%	17%

Table 4
Hindsight on foresight survey statements.

	Overall respondents (108)[a]	Chemicals respondents (11)	Energy respondents (9)	Retail & Distribution respondents (2)
Statement c) There were too many statements on each form.				
Agree	**83%**	**82%**	**89%**	**100%**
Disagree	13%	18%	11%	0%
Other	4%	0%	0%	0%
Statement f) There were too many variables on the forms (timing, UK position, constraints, etc)				
Agree	**78%**	**82%**	**78%**	**50%**
Disagree	16%	18%	22%	50%
Other	6%	0%	0%	0%
Statement d) The statements were too complicated				
Agree	**68%**	**91%**	22%	**100%**
Disagree	26%	9%	**67%**	0%
Other	6%	0%	11%	0%

[a]Overall respondents comprise respondents from the steering committee (1) as well as from all 15 panels (Agriculture , Natural resources and Environment (11); Chemicals (11); Communications (5); Construction (5); Defence and Aerospace (6); Energy (9); Financial Services (3); Food and Drink (9); Information Technology (11); Health and Life Sciences (7); Learning & Leisure (3); Materials (10); Manufacturing and Business Processes (9); Retail & Distribution (2); Transport(6)).

5.2. Realisation rates by predicted time-bands

Table 3 presents the distribution of estimates falling into the various time-bands. It is evident that the 'Chemicals' and 'Energy' estimates are more conservative, focusing on the 2000–2009 time-band, compared to the 'Retail & Distribution' estimates with 60 (75%) of statements falling into the 1995–2004 time band. However, while this "conservatism" produced better estimates in case of the 'Energy' Delphi, the 'Chemicals' survey contains the least fully as well as partially realised statements. What comes out of this data is that realisation rates decrease over time-bands, suggesting that topics take much longer to materialise as predicted. While this again is indicative of overly-optimistic estimates, it at least allows speculating that realisation rates might increase in the future as time progresses.

6. 1995 OST 'Hindsight on Foresight' survey

After the publication of the panel reports in 1995, OST sent out the 'Hindsight on Foresight' survey to all panel members collecting feedback to improve the UK Technology Foresight Programme.[5] Section four of that questionnaire dealt with the Delphi survey, asking panel members whether they 'Agree' or 'Disagree' with eleven statements, three of which are presented in Table 4. Furthermore, an open comment box was provided at the bottom of the section of which good use was made by respondents. This feedback is a crucial source of evidence to interpret the current survey results, as it indicates quality issues in the original Delphi process and format. While a total of 108 panel members participated in the questionnaire, the number of responses varied considerably between panels.

6.1. Hindsight on foresight statement results

Overall, 83% of respondents agreed that "there were too many [Delphi] statements on each form", implying that the Delphi questionnaire took too long to complete and was too broad for experts to make credible judgements. These views were also expressed as open comments. Interestingly, this view is shared across the assessed panels, where the 'Retail & Distribution' responses are generally unreliable as only two responses were made. Similarly, 78% of total respondents agreed that "there were too many variables on the form", implying that experts struggled to make reliable judgements on the variables as well as producing "shooting from the hip" judgements to save time. Again, this view is supported by open comments across panels even saying that the questionnaire was at the "limit of acceptability". This happened despite the instructions of the survey reminding experts not to feel forced to complete all statements. Furthermore, this format of 80 statements and set of variables was deliberately chosen by the sponsor and panels against the advice of the organisers of exercise (Interview, Delphi coordinator, 2006).

Interestingly, the views differ between the panels in regard to the "statements being too complicated". While 68% of total respondents agreed with this statement, the percentage is much higher for the 'Chemicals' and 'Retail & Distribution' panel, whereas more than 2/3 of the 'Energy' respondents disagreed. If statements are too complicated, experts have difficulties interpreting these, which will invariably impair the quality of the resulting estimates. In the light of the survey results presented above with 'Energy' estimates containing many more fully realised statements, this difference in opinion between panel members is evidence that differences in estimates between panels might be explained with quality issues in the process generating the statements and Delphi format.

[5] This intention by OST was made clear in the cover letter of the 'Hindsight on Foresight' questionnaire.

6.2. Open comments

The open comments made by respondents often supported the above results arguing that the statements were developed under time pressure by various people, resulting in poorly phrased and complicated Delphi statements. This happened despite rules how to phrase statements and central auditing of statements by the Delphi organisers (Interview, Delphi coordinator, 2006). However, the practice of developing the statements varied widely between panels and the sponsor decreed that the Panels could over-rule the auditors and retain their format.

> "Delphi forms are too complicated. There are too many questions, although all of them are relevant, and the forms take far too long to fill in. All of this results in a poor response or a 'firing from the hip' response, i.e. a non-considered view. Again, it is the timescale that is wrong. The forms had to be completed and sent back too quickly".

While one respondent rightly pointed out that most panel members were unfamiliar with Delphi processes, the panels were nevertheless required to design and develop the Delphi statements under the employed methodology. An unstructured first round including the whole response group to develop the statements was hence not conducted as part of this Delphi. While this is not unusual in large Delphis–Cuhls pointed out that only the first Japanese Delphi in 1971 employed an unstructured first round, whereas all proceeding surveys used expert groups to develop topic statements [1] – such a round could have been used to compensate the lack of Delphi experience in the UK. Furthermore, despite having gained 30 years of experience with the technique, the Japanese commit considerable resources and time to the development process of the Delphi statements, employing full time staff over long periods of time to ensure high quality statements.[6] While the experts in the UK developed the statements independently, some made use of the regional workshops and outside experts to "test" and improve their Delphi statements. However, as this was not part of the set methodology, this practice varied between panels. The 'Energy' panel report describes the extensive use of regional workshops to improve the Delphi statements [14], whereas the 'Chemicals' report states that 'outside' experts commented on their statements [15]. However, the 'Retail & Distribution' panel only states that statements were scrutinised internally [16]. While most panels attempted to produce the best statements possible under the circumstances, this points to quality differences in the development process between panels.

> "Generating the questions was rushed. Because of the number of questions, individual panel members produced different types/emphasis/quality questions".

Most panel members named time pressure as the most crucial constraint on the quality of statements resulting in poor phrasing and format. As Keenan pointed out the 'Health & Life Sciences' Panel for example developed all statements in a single afternoon [11]. Furthermore, the statements were often developed by sub-groups focusing on sub-sectors leading to differences in style, quality and emphasis of statements even within panels. One respondent even stated that his panel developed the Delphi statements before having formed a view about the future. These comments make an important point, questioning the quality of the Delphi statements used. However, as said above the quality seems to vary substantially between panels. This results in the expectation that the estimates based on these statements also differ between panels.

> "I think it should have [Delphi being a consultative tool], but in practice, a lot of recipients binned the form without thinking about it because it looked too complex".

Further comments of expert respondents outlined that too many statements and response variables were on the questionnaire, stretching the patience of participants resulting in at least questionable responses. Panel members reported that some participants even binned their questionnaires, not willing to participate. Although this might be a general feature of more complex surveys, this is normally not reported back. All of the above issues had impact on the Delphi statements and resulting estimates. These issues are therefore the most obvious starting point in the interpretation of the results, however, not necessarily the most important ones.

7. Interpretation of survey results

7.1. Comparison to previous Japanese Delphi assessments

So far, none of the European Delphi surveys have been analysed in retrospect in regards to the realisation of estimates, whereas the Japanese have integrated this process in their quinquennial surveys since 1996.[7] Table 5 presents the overall realisation rates of the 1971, 1976 and 1981 Japanese Delphi surveys as assessed in 2001, and compares these to the results of the assessed UK panels.

Although the identical definitions of states of realisation (Realised, Partially Realised, Not Realised) were used as in the Japanese survey, the UK assessment was conducted after only 12 years of the estimates being made, compared to at least 20 years for the Japanese assessment. Nevertheless, this data is the best comparative source available. Comparing the realisation rates clearly shows that the UK rates are much lower than the Japanese data, although the UK Energy estimates compare favourably against the

[6] A member of the 'Construction' panel pointed out on the 'Hindsight on Foresight' questionnaire that the Japanese employed several full time staff over 18 months to develop the Japanese Construction Delphi statements.

[7] The first assessment of past estimates was completed as part of the 6th Japanese Delphi survey, and repeated in the 7th Delphi survey. The results of the 8th Japanese Delphi were not fully published in English at the time of conducting this research.

Table 5
Japanese–UK Delphi comparison.

	Status assessed in 2001			Status assessed in 2006		
	Japanese (1971)	Japanese (1976)	Japanese (1981)	UK Chemicals	UK Energy	UK Retail & Distribution
Realisation rate	30%	25%	20%	5%	15%	6%
Joint realisation rate	66%	65%	71%	28%	34%	43%

Data taken from 7th Japanese Delphi survey (NISTEP, 2001).

other two panels. Furthermore, the realisation rate seems to increase over time as the 1971 survey has the highest rate of fully realised statements. This is in contrast to the 'Joint Realisation Rate', which seems to be similar across the different Japanese surveys, while this is not true for the UK results. There are two important points here. Firstly, it seems that the 'Realisation Rate' increases over time as estimates materialise much slower than anticipated. Secondly, the UK results have lower realisation rates compared to their Japanese counterparts, except for the 'Energy' estimates, which compare more favourable.

7.2. Over-optimism and behavioural aspects

Over-optimism in future estimates is nothing unusual and is well documented in the literature. Linstone and Turoff, writing about the Delphi method, highlight that short-range forecasts tend to be optimistic, whereas long-range forecasts tend to be pessimistic [5]. In the near term the solution is obvious but the difficulties of system synthesis and implementation are underestimated; in the long term no solution is apparent. This hypothesis formulated by Buschmann [17] is supported by Ament, comparing a 1964 and 1969 Delphi. He found that all items forecast in both studies originally predicted to occur in years before 1980 were later shifted further into the future, whereas two-thirds of the items originally forecast to occur after 1980, were placed in 1969 at a date earlier than that estimated in the 1964 study [18]. Consequently, Linstone and Turoff proposed to adjust forecasts accordingly e.g. to move out forecasts nearer than 10 years in time and move closer forecasts more than 10 years in the future [5]. These examples make it not surprising to find that many of the UK Delphi statements predicted to be realised by 2006 had not been realised.

However, more recent literature attempts to explain this over-optimism with experts' level of expertise. Tichy, analysing the German and Austrian Delphis found evidence that top experts tend to be even more optimistic than the overall response group [19]. Unfortunately, the analysis of the UK Delphi in regards to over-optimism of top experts is not possible as the survey data is available for the overall response group and expert group levels 3–5 only. Others argue that many future images do not come true as they are based on too simplistic conceptualisations of technological development and its relation with society, including the neglect of the dynamic co-evolution of technology and society [20].

Furthermore, at least partially offering an explanation for failed technology futures as argued by Geels and Smit [20], is the performative role of promises and expectations in the social process of technological development. Proponents of a new technology have to mobilise support and one resource for this is the promises of the new technology. For this mobilizing purpose, the advocates cannot do without some societal blinkers. Thus, it cannot be prevented that promises and expectations are part of strategic games [20]. Geels and Smit argue that on the one hand this does not constitute a serious problem as the over-optimism is fulfilling its role in the technology development process. However, on the other hand they acknowledge the seriousness of the problem when these influence policy choices including decisions on future investments. They argue that policy-makers then should not go along too easily with promises of future technologies, or in this case estimates more generally, and instead should be aware of the pitfalls. However, the question is whether policy-makers are capable of making such reflective decisions, as one of the main purposes of for example the Delphi survey is to base decisions on expert knowledge. Overruling this at a later stage seems to question the process as such. Nevertheless, this is a serious point to consider as it is not clear how future exercises could or should deal with this issue.

7.3. Specific UK programme issues

As indicated in the 'Hindsight on Foresight' evidence, the UK Delphi suffered from quality issues. There are three distinct stages in the process, which determine the quality of the statements and the resulting estimates. Firstly, the experts that develop the statements influence the content and format. Secondly, the phrasing and wording of the statements determine how these are interpreted by experts making the estimates. Thirdly, the choice of experts making the estimates as well as their attitude towards the survey determines the quality of the resulting estimates.

7.3.1. The panel experts

As indicated by the 'Hindsight on Foresight' results most panel members had no prior practical experience with the Delphi method. The main reasons why expert panels developed the statements was their expertise in the sector and the fact that the panels were supposed to represent the respective expert community [21]. While this seems to be true for the Chemicals and Energy panel, both arguing that a wide range of interest groups were represented on their panels, this seems to be less true for the 'Retail & Distribution' Panel. Being presented with the low realisation rates of the Retail & Distribution survey, one panel member interviewed stated that none of the major retailers was represented on the panel, except for one representative from 'Safeway'.

However, to guard commercial secrets from competition, this specific panel member did not contribute to the development of the statements.[8] This lack of retail experts on the panel is also admitted in the panel report, stating that the panel found it difficult to achieve an open discussion on Foresight issues for competitive reasons [16]. Furthermore, the issues raised in the foresight process did not specifically relate to the topic of distribution [16]. Due to the lack of experts directly involved in the commercial sector, it seems that the statements of the Retail & Distribution sector are largely influenced by topics dominating the media at the time and hence have a consumer focus. This view was also expressed by the interviewee, stating that the Delphi topics need to be understood in the time of making. As he argued, social cohesion was a big media topic then, resulting in statements dealing with technologies that could be used to control social tension in society. Additionally, the statements for the 'Distribution' sub-sector either deal with deliveries to the end consumer or are vague in nature. However, there are no statements explicitly dealing with technologies used in the distribution sector. This is most likely a direct result of the panel's constitution and offers one explanation why the estimates for the 'Retail & Distribution' survey have largely not been realised. Georghiou also pointed out that a significant proportion of the 'Retail & Distribution' statements dealt with legislative and other developments unrelated to the innovation cycle [12]. This is further evidence that the statements used by the 'Retail & Distribution' panel lacked the focus that is required for quality estimates. The 'Chemicals' and 'Energy' panels however, seem not to have suffered from a similar problem. As pointed out above, they were of the opinion that all interest groups from the sector were represented on their panels. This indicates that panel selection in Delphi surveys is and remains a crucial issue.

7.3.2. Delphi statements

The quality of Delphi statements influences the resulting estimates. If a statement is vaguely phrased, the resulting estimates depend on the interpretation of the statement made by the predictor. Similarly, vaguely phrased statements impede the assessment in regard to realisation. As Gordon argues: "Whether a forecast has indeed occurred is often difficult to tell, even in retrospect" [22]. Given this experience, a vaguely phrased statement amplifies this problem, as it allows various interpretations. This could serve as an explanation for the large proportion of 'Partially Realised' topics for some panels, compared to fully 'Realised' topics, as will be argued below. While no direct comments were made by the experts that assessed the statements in 2006, the substantial number of topics with 'No clear Assessment' is indicative of this issue.

However, the quality of Delphi statements used is criticised by the feedback received from panel members, although they developed the statements and were hence in control of the quality. The unfamiliarity of the method with most panel members and the fact that the use was imposed by the Programme methodology is one explanation for the criticism. The fact that the Delphi results came too late in the process to substantially influence many panel reports [13], and a resulting view that the Delphi was not very useful acts as further explanation for the motives of the criticism.

The only panel disagreeing that the Delphi statements were too complicated was the 'Energy' panel. Interestingly, this panel has a realisation rate of 24% of 1995–2004 estimates. This is close to the Japanese data compared to the 'Chemicals' and 'Retail & Distribution' results, while the 'Joint Realisation Rate' is much higher across panels. This means that significantly more topics from the 'Energy' panel were assessed as fully realised, while proportionally fewer were assessed as partially realised. In the light of the fact that the 'Energy' panel members disagreed that the statements were too complicated, the higher number of fully realised statements serves as evidence that suggests that raising the quality of Delphi statements could result in better estimates assuming 'ceteris paribus'.

Another point is the possible differences in character between statements. While there is no explicit reference in the OST reports stating that the Delphi statements were designed as guideposts, as one member from the 'Chemicals' panel put it, the 'Chemicals' statements were a "balance between expectations of what would happen in particular time scales, and a wish-list of target areas that were identified as important for the research community to work on".[9] Furthermore, as he put it, "these were challenging topics to direct the thinking of people working in the area, which could be achieved given that people focused their minds on these particular activities". This indicates that at least some statements were designed to challenge the research community and hence be difficult to put into practice. Nevertheless, this only partially explains why such statements were predicted to come about fairly quickly despite their challenging nature.

7.3.3. Delphi respondents

Another explanation for the low realisation rates can be found in the circumstances under which participants made estimates. As pointed out by panel members, it was felt that the questionnaires contained too many statements and response variables making it difficult for respondents to complete, despite the Delphi guidelines being explicit that respondents only had to answer those statement where they had expertise. This not only resulted in lower response rates, but also few experts being able to assess all statements. This indicates both a process and behavioural aspect with this Delphi. While the 'Energy' and 'Chemicals' panels generated response rates around 30%, which are comparable to similar Foresight exercises throughout Japan and Europe, i.e. Germany, Austria, France, the Retail & Distribution panel only received 67 (16%) responses. The estimates, at least for the 'Retail & Distribution' panel, are therefore based on a relatively small number of experts. Furthermore, because few respondents were able to assess all statements, respondents sometimes doubted the results of the survey. If they felt uncomfortable assessing statements due to lack of knowledge, they felt other experts must have made similar assessments [23].

[8] As revealed in an interview with a member of the 'Retail & Distribution' panel.
[9] This view was expressed by the interviewee from the 'Chemicals' panel.

8. Conclusion

Acknowledging the limitations of the research in terms of scope, it was nevertheless revealed that only a handful of the 1995 UK Delphi estimates were put into practice over the last decade. Although the Japanese results suggest that this number will continue to increase over time as it takes much longer for estimates to materialise, this is further evidence supporting the hypothesis of Linstone and Turoff that short-range forecast tend to be optimistic. Several explanations were explored to explain these low realisation rates.

The short-range optimism and strategic gaming explanation was complemented with exercise specific factors. Process issues were outlined at every stage in the Delphi process, starting with the selection of experts for the 'Retail & Distribution' panel, the quality of Delphi statements across panels in regards to format, phrasing and wording as well as the lack of practical experience with the method in the UK more broadly and the panel members specifically. In comparison the Japanese after using the method for several decades now have a well honed group of expert respondents, which is likely to improve the results of the survey possibly reflected in the higher Japanese realisation rates. Furthermore, participants complained about the lengthy and complicated format, feeling unable to make informed judgements on all issues as well as producing "firing from the hip" estimates, despite not being required to answer all statements. While the process issues are specific to this exercise and might be difficult to eliminate in future, as all exercises of this scale are very context dependent, at least the issue of short-range optimism and behavioural aspects of experts making forecasts should be addressed more closely.

While Geels and Smit do not see over-optimism and strategic gaming of experts per se as a serious problem, it very well becomes a problem when these feature in decision making and investment decisions. This at least applies to the UK Foresight exercise. They argue that policy-makers should take this factor into account during the decision making process. However, this asks a lot from policy-makers that rely on expert advice. Instead, it should rather be thought about how to deal with short-range optimism and strategic gaming of experts in future exercises and how to possibly formalise this process. Just acknowledging these results as a further piece in the overall puzzle of over-optimism in futures research is at least not sufficient. One idea could be a 'correction factor' that is applied to forecasts to improve accuracy. This correction factor could be a function of time reflecting the bias of short-range and long-range forecasts as suggested by Linstone and Turoff, but also take into account the participants' expertise related to the over-optimism hypothesis of top-experts.

But as attractive such a correction factor may sound it also has severe limitations. Firstly, such a correction factor would raise ethical issues of overruling expert estimates and puts much faith in the people making the corrections. Additionally, this would put experts that in the first place take this bias into account at a disadvantage as their estimates are 'double corrected'. Lastly, while a correction factor might solve unintended optimism, the factor of strategic gaming cannot be resolved as experts falling in this category could adjust their estimates in light of the expected correction factor leading to even more optimistic forecasts. In any case, for this to be implemented it is important to understand how foresight results are used in the decision making process and hence how better accuracy would lead to improved decision making. This should be the ultimate goal of any improvement and should be further explored in future research. Additionally, users of Delphi surveys should very carefully think about *when* employing Delphi surveys as a useful source of expert estimates for decision making processes and *how* to deal with the issues highlighted in this paper.

Acknowledgements

I would like to thank Michael Keenan for supervising the MSc research project this article is based on as well as for the encouragement and support to write this article. I would also like to thank Denis Loveridge for his many comments and help in writing this article, as well as for co-supervising the MSc research project. I am also indebted to the anonymous referees for their important and valuable work.

References

[1] K. Cuhls, Foresight with Delphi surveys in Japan, Technol. Anal. Strateg. Manag. 13 (4) (2001) 555–569.
[2] H. Grupp, H.A. Linstone, National technology foresight activities around the globe: resurrection and new paradigms, Technol. Forecast. Soc. Change 60 (1999) 85–94.
[3] D. Loveridge, L. Georghiou, M. Nedeva. United Kingdom Technology Foresight Programme: Delphi Survey" A Report to the Office of Science and Technology, PREST, University of Manchester, ISBN 0 946007 06 3, 1995.
[4] K. Cuhls, Foresight / Vorausschau: Delphi Studien zur globalen Entwicklung in Wissenschaft und Technik, in Trendforschung – Die Märkte von morgen entdecken, IHK, Nordreihn-Westfalen, 2006.
[5] H.A. Linstone, M. Turoff, The Delphi Method: Techniques and Applications, Addison-Wesley, Massachusetts, 1975.
[6] F. Brandes, The UK Technology Foresight Programme: An Assessment of Expert Estimates in Delphi Surveys, unpublished MSc Thesis, University of Manchester, 2006.
[7] O. Kuusi, Modern Delphi Method, VATT, Parliament of Finland, (10.08.2004).
[8] M. Turoff, Design of a policy Delphi, Technol. Forecast. Soc. Change 2 (1970) 149–171.
[9] T.J. Gordon, O. Helmer, Report on a Long Range Forecasting Study, RAND P2982, September 1964.
[10] K. Cuhls, Japanese S + T Foresight 2035, The European Foresight Monitoring Network: Foresight Brief No. 035, [Website] Available from: http://www.efmn.info/kb/efmn-brief35.pdf [Accessed: 24.07.2006].
[11] M. Keenan, An Evaluation of the Implementation of the UK Technology Foresight Programme, PhD thesis, University of Manchester, 2000.
[12] L. Georghiou, The UK technology foresight programme, Futures 28 (4) (1996) 359–377.
[13] L. Georghiou, M. Keenan, Role and Effects of Foresight in the United Kingdom, Paper for Workshop on EU-Enlargement: "Working Group on Thematic Network on Foresight Activities", Nicosia, Cyprus, April 7th–9th, 2000.

[14] OST, Progress Through Partnership—Report from the Energy Panel of the Technology Foresight Programme, HMSO, London, 1995.
[15] OST, Progress Through Partnership—Report from the Chemicals Panel of the Technology Foresight Programme, HMSO, London, 1995.
[16] OST, Progress Through Partnership—Report from the Retail & Distribution Panel of the Technology Foresight Programme, HMSO, London, 1995 (4 pp.).
[17] R. Buschmann, Balanced grand-scale forecasting, Technol. Forecast. 1 (1969) 221.
[18] R.H. Ament, Comparison of Delphi forecasting studies in 1964 and 1969, Futures (1970) 43.
[19] G. Tichy, The over-optimism among experts in assessment and foresight, Technol. Forecast. Soc. Change 71 (2004) 341–363.
[20] F.W. Geels, W.A. Smit, Failed Technology Futures: Pitfalls and Lessons from a Historical Survey, Futures 32 (2000) 867–885.
[21] UK Technology Foresight Programme methodology.
[22] T.J. Gordon, Chapter 3: The Delphi Method, In Futures Research Methodology – V2.0, AC/UNU Millennium Project, 2003, p.15 [Available online]: http://mp.cim3.net/file/project/mp-sofi-sd/reference/03-Delphi.PDF [Accessed: 27.04.2006].
[23] OST, 'Hindsight on Foresight' survey.

Felix Brandes holds a degree in Science and Technology Policy and Management with a distinction from PREST, Manchester University as well as a degree in International Business, Finance and Economics. His final thesis analysed the predictions of the Delphi survey conducted as part of the UK Technology Foresight Programme in 1996. Since 2006 he works as a researcher at the Innovation Policy Group at TNO, Netherlands. His work focuses on issues related to the future competitiveness of manufacturing taking into account factors of technological and social change.