

AI.1 The uneasy relationship between scenarios and prediction

To complement chapter 2, which searched the theoretical literature of futures studies and from that deriving some criteria for evaluating a method, this paper takes an empirical approach, comparing a group of published scenarios for the year 2000 with the outcomes in that year. Even though scenarios are not intended to be predictions, by making such a comparison, it was expected that some problems with the practice of foresight might emerge that were not evident from the previous chapter.

Though modern scenario planning disavows any pretence of prediction, Bell and Olick (1989) argue that some form of prediction is inherent, or implied, in any futures studies exercise. Among scenario developers it is recognized that no scenario will be accurate, no matter how large the ensemble (van der Heijden, 1996; Schwartz, 1991; Slaughter 1995a). However, the clients of futurists may reasonably hope to find the eventual outcome lying somewhere in the dimensional space filled by the scenarios. In fact, the currently most common method of scenario building, the “critical uncertainties” approach, explicitly searches for the two or three most critical dimensions on which the future might vary, by first creating a hypercube of that number of dimensions (i.e. a square for two dimensions, or a cube for three dimensions). The next step is to create an ensemble of scenarios that correspond to the vertices of the hypercube. If those dimensions are well chosen, the outcome must, by definition, lie somewhere within that hypercube.

Though single scenario can ever expect to fully predict an outcome, if a study has examined a comprehensive scenario space, it should be possible to locate the eventual outcome somewhere in that scenario envelope. The most useful results are likely to be achieved when the outcome, as it materializes, reveals a position similar to one of the scenarios. In that case, the users of the scenarios can point to a metaphorical map, saying “now we are here.” If they have already identified the likely consequences of that position, they will be in a position to take appropriate action. Though this approach may be controversial among the advocates of the

1. An earlier version of this paper was presented at the conference “Innovation and Social Development in the Knowledge-Based Economy/Society,” Dubrovnik, Croatia, May 2004 (List, 2004e).

“scenario learning” position (such as van der Heijden, 1996), it is supported by Slaughter, one of the most respected futurists of the present time:

It is logically barred to us that we could ever have future facts about human and cultural systems. So we move to the next best option. That is a provisional, but fairly detailed and grounded picture, or view, of the terrain ahead. This view is continually informed and updated as events occur and our interpretations of the world change. (Slaughter 1995a:32)

In fact, the above passage (like the next, from the same book) can be interpreted not only as a justification for the approach taken here, but also as a broad description of scenario network mapping:

In general terms, the purpose of futures work is not, as I have indicated, to predict. It is rather to ‘map’ the near-term future insofar as this is possible with the knowledge of, and from the vantage point of a particular point in time. Like any map, including maps of the physical world, there will be much that cannot be represented. But the futures map does try to indicate such things as areas of danger, areas needing attention and a range of directions, options and alternatives. Given that people have had so much practical experience in foresight, no-one should have too much trouble ‘reading’ this futures map. (Slaughter 1995a:51)

However, this is not prediction in the generally accepted sense of the term (as critiqued in chapter 3). A more appropriate word, used in this study, is foresightfulness: it implies a level of foresight which turns out to be justified by later developments.

A1.2 Basis of this analysis

Bearing in mind Slaughter’s and Bell’s views, and my initial hope that SNM might enable scenario users to place their entity somewhere on the metaphorical territory of a map of the future, it seemed useful to compare the “predictions” of various foresighting techniques (particularly scenario planning) to outcomes. The aim of this content analysis was to develop an understanding of the problems involved with the development of foresighting, so as to be able to find some ways of overcoming them.

In order to assess foresightfulness, this study sought anticipations of a wide range of types (particularly scenario ensembles), which met all of the following criteria:

1. extending over an outlook period extending from at least 5 years ahead (because a shorter time span would make the assessment of outcomes more ambiguous);
2. whose target dates had already passed;
3. for which the outcomes data would already be available.

Because much had been written anticipating the millennium year 2000, and also because it had recently passed, I chose that year as the target date, and searched for published anticipations to review in this process. (The word “anticipations” is used here to refer to any kind of description of a possible future, including but not limited to scenarios; the word “foresight” here has a similarly broad implication.) The specific questions addressed were:

- For each document: to what extent did the outcomes match the anticipations?
- How did the degree of matching vary according to the scale and type of foresighting method used? And what can this tell us about the effectiveness of each foresighting method?
- To what extent did the matching increase with proximity to the target date? Specifically, were the later anticipations closer to outcomes than the earlier ones, and was there any point at which the outcome suddenly became generally “correct”?
- What general problems became evident on reviewing these anticipations?

Answers to these questions are addressed in sections A1.4 to A1.7 below, and section A1.8 offers some lessons for a futures Process.

A1.3 Summary of studies

I obtained abstracts, reviews, or full documentation for 128 anticipations that had been published before the main cut-off date of 1995 (or 1997 for the ICT industry, which changes rapidly). The following table shows the number of works for which some details were found, and the numbers used in this review.

TABLE A1.1 SUMMARY OF ANTICIPATIONS OF 2000

Type of work	Found	Used
Single scenarios, developed in detail - “fiction,” utopias, intuitive reports: De la Bretonne (1771), Bellamy (1888), Morris (1890), Kipling (1905), De Closets (1978)	6	2
Scenario ensembles, created in formal scenario planning	64	7
Sample surveys: Gallup and Proctor (1984), Ornauer et al (1976)	8	2
Other types of scenario and anticipation (mainly econometrically based)	22	3
Anthologies, describing a range of other works: Armytage (1968), Corn and Horrigan (1984), Walter (1992), Canto and Faliu (1993), Jencks (2000)	5	1
Uncertain - have not found enough detail to determine	23	0
Total	128	15

Unfortunately, many of the 128 studies could not be evaluated in detail, for one or more of the following reasons:

- (a) The full report was unobtainable;
- (b) The report was in a language that I could not read;
- (c) Outcomes could not be evaluated, for lack of data on the situation in 2000;
- (d) The study provided insufficient detail for outcomes to be evaluated;
- (e) The study turned out to use only econometric or mathematically derived scenarios on a small range of variables, rather than scenario planning. This was the commonest reason for exclusion.

After allowing for those exceptions, the final sample was chosen on a purposive basis. That purpose was to collect documents from a wide range of different anticipatory approaches (but particularly scenarios), in order to both compare the effectiveness of these different approaches and to discover problems that might be overcome with a new method. Documents with later published reviews were favoured. The final sample included 15 studies. The following table describes these, using the following headings:

Name of study:	A brief name, used to refer to this study throughout this document..
Reference:	The major published document on this study.
Base year:	The year the study was performed or completed. When no information was supplied in the document, the year before publication appears here, italicized.
Area scope.	When this is not explicitly stated in the document, the country or area of origin is italicized.
Subject scope:	The content focus of the study.
Method:	Distinguishes between intuitive (e.g. science fiction writers) and expert intuitive (an intuitive forecast, informed by extensive domain knowledge).

Studies in the table are arranged in a sequence from largest to smallest scope: firstly, general, worldwide studies, then general country-based, then industry-based, and finally studies of a single industry in one country. The discussion below the table follows the same sequence.

TABLE A1.2 ANTICIPATIONS OF 2000 INCLUDED IN THIS REVIEW

Name of study	Reference	Base year	Area scope	Subject scope	Method
The Year 2000	Kahn & Wiener (1967)	1966	world	general	1 main chain scenario, with derivatives; also forecasts and modelling
Mankind 2000	Jungk & Galtung (1969), Ornauer et.al (1976), Galtung (2003a, 2003b)	1967	10 countries, rich and poor	general	Surveys of experts and general public
OECD Interfutures	Norse (1979)	1976	world	International economic development	Scenario planning by team of experts
Global 2000 (Carter Report)	Barney, 1980	1977	world	environmental, economic	Econometric modelling of official statistics
Perspective 2000	Newton et.al (1990)	1988	world	political, economic, environmental	Economic: 3 scenarios, (critical uncertainties)
Forecast 2000	Gallup & Proctor (1984)	1984	USA		Surveys (experts and youth)
Australia at the Cross-roads	Kasper et.al (1980), Galer & Kasper (1982)	1980	Australia	macro-economic	2 scenarios (Shell method)
Today Then	Walter (1992)	1893	USA	anthology: comprehensive	Expert intuitive
Norway 2000	Osmundsen (1986), Nore & Osmundsen (1988)	1987	Norway	Comprehensive	Scenario planning (3 scenarios)
Poland in 2000	Wierzbicki (1991)	1985-89	Poland		Scenario planning (3 scenarios)
Europe 2000	Hall (1977)	1974-76	Europe	Comprehensive, etc	Team of experts
Information commerce	Randall (1997)	1996	USA	Business use of the internet	Critical Uncertainties scenarios, with drivers
As We May Think	Bush (1945)	1945	USA	Information technology	Expert intuitive
Architecture 2000	Jencks (1971, 2000)	1968	world	Architecture and related areas	Expert intuitive
Telecom 2000	Telecom Australia (1975)	1975	Australia	Telephone communications	Mixed methods, no scenarios

Each of the 15 studies is now compared on the five research questions raised in section A1.2 above.

AI.4 To what extent did the outcomes match the anticipations?

On this first issue of inquiry, each study is compared separately, beginning with a brief summary of the study.

Study 1. Global 2000 (Barney, 1980)

Topic:	The natural world as it could be in 2000.
Method:	Computer analysis, using a modified version of the econometric model developed for the Club of Rome, and described in Meadows et al. (1970) and Mesarovic and Pestel (1972).
Output:	A set of numerical parameters of major economic, social, and environmental measures.

This generally pessimistic book was followed by *The Resourceful Earth*, an optimistic riposte produced by Simon and Kahn (1984); outcomes generally fell between the two. Chenoweth and Feitelson (2005) compare *Global 2000* and *The Resourceful Earth* in the context of a debate between Neo-Malthusians (the former) and Cornucopians (the latter), concluding that “as a whole *The Resourceful Earth* was more accurate”. Cole (2005) somewhat disagrees with Chenoweth and Feitelson’s conclusions, emphasizing their commonalities, taking a similar approach to that of Richardson (1985). Dator (2005), also commenting on Chenoweth and Feitelson, also disagrees with their conclusion, and points out that both books were intended more as political influences than as scientific prediction. My own view is that on many issues (such as world population) both books produced broadly accurate projections, but neither anticipated the widening gap between different groups of countries - such as life expectancy rising in rich countries while falling in Africa, perhaps because at the time economists were generally focused on aggregates rather than on distributions; cf. Atkinson (1982).

Study 2. The Year 2000 (Kahn and Wiener, 1967)

Topic:	The world (on a macro scale) in 2000.
Method:	Eclectic one, incorporating forecasts and projections as well as scenarios, which appear to have been intuitively derived.
Output:	One main “surprise-free” scenario, with a set of eight “canonical variations”, in three groups: more integrated worlds, more inward-looking worlds, and worlds with greater disarray.

The more general anticipations were expressed so vaguely that it is not possible to say precisely whether or not they occurred: for example, “the rise of new great powers” (could that include the EC or China?), “new political, perhaps even ‘philosophical’ issues” (to what extent can an issue be described as “new”, versus a refocusing of an old issue?). Clearer anticipations were often not borne out: there are no indications that communism might end, and the rapid economic growth in China that occurred in the 1980s and 1990s was regarded as unlikely. Of a list of 100 technical innovations, 30 were clearly in use by 2000, 50 were not in use, and the other 20 were uncertain.

Some of the outcomes foreseen in a later but similar book by Kahn (1976) were reviewed by Galtung (2003b). On comparing 11 of Kahn’s predictions for Western Europe in 2000 with the actuality, Galtung found only 7 to be substantially wrong, and 4 to be partially wrong, including the “normalization of both Germanies.” As Galtung suggests, Kahn had the US government as his major audience, and the findings may have been coloured by that.

Study 3. Interfutures (Norse, 1979)

Topic:	The economic prospects of the world in 2000,
Method:	Economic modelling was used, but as an input to the scenario planning process, not as an output in its own right.
Output:	Six end-state scenarios, with a limited analysis of dynamic paths leading to some of the end-states defined. Three scenarios were slight variations on a theme. The report also produced a “mixture” scenario, preferred as being most plausible.

The “mixture” scenario (though lacking specific numbers) is closest to the economic climate of 2000. A remarkable characteristic of these scenarios is that, in a supposedly world-wide study, they completely ignore the “Communist bloc” (as it was then). Thus its dissolution is simply not envisaged. Though almost no other writers envisaged a dissolution, except Shell Oil’s scenario group (Bennett and Khalifa, 2000), and to some extent Ayres (1979, pp37-38), with a retrospective claim by Dewar (2002, p190), the failure of this study to even consider the Communist bloc must be regarded as a weakness - particularly given its proximity to Western Europe, which was the focus of the study, and the fact that the study did include other continents.

Study 4. Mankind 2000 (Ornauer et.al, 1976)

Topic: Images of the world in the year 2000, among the general public and experts.
Method: Surveys of 11,000 people in 10 countries, and of experts.
Output: Responses to individual survey questions.

In a follow-up report to this study (Galtung, 2003a and 2003b) the findings of the 1967 survey are contrasted with the experts' prognostications for 2000, and concludes that while the general public in 10 countries (in aggregate) had fairly accurate expectations of the future, those of the experts (discussed in detail in Galtung, 2003b) were generally wrong, because those experts "see the world from above, in a system perspective, from the vantage point of the rich and powerful." He concludes that "the most relevant question to an expert is not *what do you see*, but *who paid you?*" These remarks are reinforced by the generally poor performance of experts in studies 7 and 8 (below).

Study 5. Perspective 2000 (Newton et.al, 1990)

Topic: Economic futures for G8 countries in 2000, from a Canadian viewpoint.
Method: Three scenarios were developed by experts.
Output: Scenario 1. low openness, low co-operation: i.e. revival of protectionism
Scenario 2. high openness in international trade, without government economic co-operation: "fluctuations" (seen as most probable)
Scenario 3. high openness, and high co-operation (the most desirable for Canada).

It was clear that in 2000 the G8 economy was clearly (as expected) still in Scenario 2. However, economic growth in China was far above the predicted rate - even in the late 1980s, while Japan's growth was far below the rate expected in Scenario 2. This revealed a problem with these scenarios: the conflation of co-operation and economic openness across all countries: the scenarios did not deal with the possibility that not all economies would fall into the same quadrant. So on the whole, even though the (G8) world as a whole lay in the predicted quadrant in 2000, many of the implications that grew from that scenario were not fulfilled.

Study 6. Europe 2000 (Hall, 1977)

Topic: The futures (in several aspects) of Europe in 2000.
Method: Very large scale study, using full-time workforce of experts.
Output: Five scenarios.

Outcomes are clear: scenario 2 applies (“Western Europe united, with a politically strong EC”) and so does scenario 5 (“Explosive change in Eastern Europe”). The latter was dismissed by the authors as “prospects...almost non-existent,” despite precedents such as the Prague Spring of 1968). The brief treatment of scenario 5 may have been added as an afterthought, as if the 200 experts noticed at the last minute that Eastern Europe had so far been ignored. The actual outcome, of a strong EC extending into Eastern Europe (at least with “candidate states” in 2000) would have been even more unthinkable than scenario 5. In fact, in the remainder of the book, Eastern Europe and the USSR are mentioned only in passing, on a few pages - yet nowhere is it stated that the scope of the study was to be confined to Western Europe.

Study 7. Today Then (Walter, 1992)

Topic: The USA in 1992, as foreseen from 1892.

Method: 74 contributors were asked to state their expectations for 1992, addressing 33 questions.

Output: Given the limitations of the format, most writers devoted only a sentence or two to each question they addressed, and many did not address all the questions.

In the same way that the questions reflected the times (e.g. those quoted above on the servant problem, temperance, and suffrage) so did the responses. Some of the keenest perceptions came from those who were well-informed, but without vested interests. Thus Thomas Dixon Jr, a lawyer, legislator, and Baptist minister, foresaw a wide range of changes that would occur well into the 20th century - both social and technological, including the increasing complexity of society. On the whole, the record of these experts was poor - as with Galtung’s study. For example, though the automobile already existed in the 1890s, and one question specifically asked about the future of transport, not one of the 74 mentioned the widespread use of personal motor vehicles in the 1990s.

Study 8. Forecast 2000 (Gallup and Proctor, 1984)

Topic: Life in the USA in 2000.

Method: Survey of 1,346 opinion leaders in the USA.

Output: A book containing the survey results and discussion of them.

Given the very general nature of the survey questions, outcomes are difficult to compare with these anticipations, but it is clear that the experts were quite wrong on many small details. This book is that it is prescriptive rather than forward looking: it describes its own time much

better than it describes 2000: what the experts regarded as desirable, perhaps, rather than what they expected to happen.

Study 9. Poland in 2000 (Wierzbicki, 1991)

Topic: Macro-scale issues in Poland for 2000.
Method: Large scale study with full time staff.
Output: Three normative scenarios, with priorities for (1) energy and resources; (2) consumer demand, and (3) “civilizational progress.”

On comparing the three scenarios from this study with the Economist Intelligence Unit’s 2001 *Country Profile for Poland*, (Economist Intelligence Unit, 2001b) the outcome for Poland in 2000 appears to lie between Scenario 2 and Scenario 3, but closer to Scenario 2. Some changes that were expected to arise from Scenario 3, such as an increase in tertiary education levels and a population flow from villages to cities, were well under way in 2000, with the participation rate in higher education rising from 10% in the 1980s to 28% in 1999-2000. The contribution of heavy industry to the Polish economy had fallen rapidly, with increased concerns about pollution and joining the EU. Emissions of dust and of sulphur dioxide approximately halved between 1990 and 1999, partly reflecting the closure of much of the heavy industry that caused the pollution. Given the massive changes arising with the end of Communism in Poland, this study performed well in anticipating the future of the country - except for the political future.

Study 10. Australia at the Crossroads (Kasper, Blandy, et.al, 1980)

Topic: The Australian economy in 2000.
Method: Large scale study with full time staff.
Output: Econometric analysis, from which two scenarios were intuitively derived.

The outcome in this case is very clear: Australia moved decisively toward the “libertarian alternative” in the mid-1980s, with extensive economic deregulation. However, the scenarios were for 2000, not the mid-1980s, and some of the supporting data - derived from forecasting in the two scenario situations - were very wrong indeed. Far from expanding the manufacturing sector at the expense of the service sector, the projected sectoral share of manufacturing in 2000 turned out to be one third less than in 1990.

This study was very likely a contributing factor to the economic policies adopted by the new Labor government after the 1983 election. The presentation of two scenarios, with the

“business as usual” model widely perceived in official quarters as dysfunctional, caused governmental attention to fall on the libertarian scenario, which in turn perhaps eased its accomplishment.

Study 11. Norway 2000 (Nore and Osmundsen, 1988)

Topic:	The Norwegian economy and society in 2000
Method:	Large scale study with full time staff.
Output:	Three scenarios.

These scenarios were compared with outcomes, based on the 2001 profile of the Norwegian economy from the Economist Intelligence Unit (2001a). Of the three scenarios, one was clearly dominant; it foresaw continuing expansion of the welfare state, more influence for women, slower economic growth, and Norway still not being a member of the EC. All of these outcomes occurred. This was an interesting example of a set of scenarios with a background agenda of change, which socially conservative forces prevented from happening. In one sense the outcome of the scenario planning was successful, in that one of the three scenarios was essentially correct. In another sense, the scenario planning was unsuccessful, in that there had been no change; this makes a clear contrast with the Australian scenarios discussed in Study 10.

Study 12. Architecture 2000 (Jencks, 1971)

Topic:	Architecture in developed countries in 2000.
Method:	Intuitive foresight by one expert
Output:	60 predictions.

In 2000, Jencks followed up his 1967 book with a new book, following up the earlier work. In retrospect, he scored his predictions at 37 right, 15 half-right, and 8 wrong - though perhaps an independent assessment may not have been so kind. By my own reckoning, 30 of the 60 predictions were clearly correct in 2000. As many of the predictive statements were not at all “safe” (e.g. the rise of complexity theory and of the post-modern/ neoclassical style of architecture, neither of which existed in 1971) this is a highly successful achievement, perhaps arising through Jencks’ thorough immersion in his subject. As an independent commentator rather than a practitioner, he was not in a position to gain from the fulfilment of his predictions.

Study 13. As We May Think (Bush, 1945)

Topic:	Technology for knowledge management, but not specifically for 2000.
Method:	Personal intuition, from Dr Vannevar Bush
Output:	An article in the <i>Atlantic Monthly</i> .

In this article, Bush describes devices that later became the fax machine, digital cameras, hyperlinking, and the “memex” - almost a modern PC, including a mouse with a scrolling wheel), multiple screen windows, and several database-like features that do not yet exist. On the whole, the degree of foresight was remarkable. Since Bush was a chief scientist for the USA during World War II, he was probably informed by a wide range of other scientists, so this article could well be a synthesis of available data rather than the result of pure intuition.

Study 14. Internet commerce (Randall, 1997)

Topic:	Commerce on the internet in 2000
Method:	The “critical uncertainties” scenario method, interviewing 30 leading experts.
Output:	Four scenarios.

Comparing data for 2000 with Randall’s four scenarios, the only scenario not fully borne out was “Crumbling walls.” To a large extent, all four scenarios were true in 2000 - for the Worldwide Web. However the most frequently used facility on the Internet (email) was not even mentioned in this paper, which also failed to anticipate the rapid growth of business-to-business e-commerce. The latter was outside Randall’s declared scope, but one might have hoped that 30 leading experts might have drawn his attention to these factors.

Postma and Liebl (2005) made similar comments on Randall’s paper: “if a company had constructed these four scenarios and had been focusing on the drivers without realizing the segmentation, it might have become helplessly confused when interpreting the occurring events because the level of analysis (the Internet) turned out to be inappropriate” (Postma and Liebl, 2004, p8).

Study 15. Telecom 2000 (Telecom Australia, 1975)

Topic:	Telecommunications in Australia in 2000
Method:	Large scale study, mostly using experts, with many external consultants.
Output:	A mixture of forecasts and papers; no formal scenarios.

This report offers a clear demonstration that what is measurable may not always continue to be relevant: its prediction of the percentage of Australian households with a telephone in 2000 was accurate, but it failed to foresee the growth of mobile telephony. Nor did this study foresee the huge changes in the industry that occurred in the 1990s, in terms of ownership, scale, and customers' willingness to spend far more money on telecommunications than previously.

However, as Ascher (1981) points out, the most accurate forecasts are those made by authorities that can regulate both supply and demand. Telecom Australia, as monopoly supplier of telecommunications services until the late 1990s had exactly this characteristic. With high latent demand for telephone services in Australia, Telecom Australia, by controlling its rate of connections, was free to fulfil its own forecasts. Had it not been the introduction of competition in the 1990s, the forecasts in *Telecom 2000* might have been even more accurate.

Telecom Australia (1978) followed up the 1975 publication with a report on its outcomes. Chapter 4 deals with reactions to the original report, including those from Telecom's own staff. Many responses criticized the emphasis and assumptions in the 1975 report, but only one – from a media organization – was recorded as pointing out the focus on the organization itself rather than its context.

A1.5 How did foresightfulness vary with the foresighting method used?

For each of the 15 studies, the following table considers the level of foresightfulness, operationally defined here as the difference between anticipations and outcomes. The column headings are interpreted as follows:

Method. This column summarizes the method/s of anticipation used in the study.

Scale: shows the scale of the work, as an estimated number of person-years. As this factor was often not reported, many of these entries are necessarily rough estimates; such entries are followed by a tilde. The person-years figure does not include time spent in data collec-

tion, for either respondents or interviewers; thus the focus is more on the time spent in high-level thinking.

Creators: the category of persons who produced the scenarios or other anticipations: whether experts in the subject matter (for these studies, most were economists), futurists, or members of the general public. “Experts” may include some futurists; not all documents specified this.

Match: a comparative five-point scale (denoted in a semi-graphical form by using between one and five bullet symbols) to rate the difference (in my assessment) between anticipations and outcomes. The comparison is not of course absolute, but reflects a broad ranking of the studies. Thus a single asterisk means “among the lowest level of matching of all this sample” while five asterisks denote “among the highest degree of matching among this sample.”

TABLE A1.3 DIFFERENCE BETWEEN ANTICIPATIONS AND OUTCOMES

Study	Method	Scale	Creators	Match
Mankind 2000	survey in 10 varied countries	5	11,000 public	*****
Architecture 2000	intuitive forecasts by one expert	1	1 expert	****
As We May Think	1 intuitive scenario	<1	1 expert (+advisers?)	****
Australia at the Crossroads	2 scenarios: status quo & normative	5	experts, futurist	****
Perspective 2000 (economic)	3 scenarios: critical uncertainties	1~	2 experts	***
Norway 2000	3 scenarios: normative	15~	experts	***
Poland 2000	3 scenarios: normative	10~	experts	***
Telecom 2000	modelling, forecasting	10~	experts	**
Europe 2000	5 scenarios: mixed methods	200~	experts	**
Kahn & Wiener	1 main + 8 other scenarios, intuitive; forecasting	8~	futurists	**
Interfutures	6 scenarios: mixed methods	10~	experts	**
Internet commerce	4 scenarios: critical uncertainties	1~	futurist	**
Forecast 2000 (Gallup)	survey of 1,346 experts	2~	1,346 famous people	**
Global 2000	econometric modelling	20~	experts	**
Today Then	survey of 74 experts	2	74 famous people	*

Before discussing the implications of the above table, there are three types of reservation to be taken into account: the possible effects of selection bias, limitations of the outcome

variable, and the possibility of self-fulfilling prophecies: anticipations influencing their own outcomes.

AI.5.1 Possible effects of selection bias

A commonly noted problem with assessing predictions is that the less successful predictions tend to be quietly forgotten (Sherden, 1998). Therefore, a search of the published literature is likely to under-represent forecasts that were well off the mark. For example, Vannevar Bush (1945) is much cited, perhaps because of his prescient image of the Internet. Acknowledging that this affirmative bias may exist among the documents found for this project, the issue is to what extent it might differentially affect the different types of method. For example, were the more successful individual predictions (such as Bush's) more easily found than the more successful scenario plans, because of the numerous references to the former? For every Vannevar Bush, could there be a dozen Mother Shiptons?²

A strong argument against such bias having occurred in this study is its thorough sampling of original documents, with only the Bush paper sourced from post-2000 citations.

AI.5.2 Limitations of the outcome variable

The Matching scale, the outcome variable in the above table, has several limitations:

Selection of statements for matching. Because many of the documents were entire books, and most of these included hundreds of anticipatory statements, selection of one subset of statements rather than another might produce quite different ratings for outcome matching. To minimize the effect of my own subjectivity, I relied as much as possible on the authors' own summaries.

Subjective matching. However, the assessment of the matching between the anticipation and the outcome is mostly my own. I originally intended to recruit a group of judges for this task, but inquiries failed to find anybody willing to devote perhaps a full month to reading all fifteen documents being studied, as well as the comparative outcomes data. I attempted to enter this matching process without prejudgements: the purpose here was not to hypothesize that one type of foresighting method produces more "accurate" anticipations

2. Mother Shipton was an English prophet of the 15th century, whose name was attached in Victorian England (Harrison, 1881) to rhymes such as

"In eighteen hundred and eighty one
The end of the world shall surely come."

than another, but rather to unearth the more subtle factors that made some anticipations closer than others. Though it is possible that another judge might produce quite different results on the Matching score, I presented tentative findings at both a local seminar and an international conference – to audiences predominantly of economists rather than futurists. Disagreement was expressed with only one judgement in the above table, and then only mildly. Further, writers who have made similar comparisons of anticipations and outcomes (including Wise 1976, Cornish 1997, Galtung 2003b, and Postma and Liebl, 2004) have drawn conclusions similar to my own.

Given the above reservations, the scoring method was restrained to a simple 5-point scale: finer judgement could have been spurious.

A1.5.3 Could anticipations have influenced their own outcomes?

To some degree, each of the above anticipations could be a self-fulfilling prophecy, or self-negating – the “Oedipus effect” of Popper (1957, p13) or the “bandwagon effect” of Henshel (1982). On the whole, however, the potential influence of most of these studies was low, and none of the above studies could be described as self-negating. However some of the most widely publicized studies were potentially self-fulfilling. This is partly because these studies were (necessarily) published, and published studies tend to describe large entities – such as entire continents. Thus the sponsors of such studies have less scope to influence the outcomes than does (say) an oligopolistic business to influence its own future. The highest influence potential occurs when, for example, a study of a country’s future is commissioned by the government of that country, *and* one of those futures is officially preferred, *and* that government remains in power until that preferred future is implemented. This applies to three studies in the above list: *Norway 2000*, *Australia at the Crossroads*, and *Telecom 2000*. For these, the matching score was correspondingly adjusted downwards by one step.

A1.5.4 Comparing anticipations with outcomes

Bearing in mind the above limitations to drawing firm conclusions from Table A1.3, the columns headed Method, Scale, and Creators permit three kinds of tentative conclusions:

In terms of the anticipatory method. In general, the closest matches occurred for intuitive and normative scenarios. (As the Ornauer survey was *sui generis*, no clear conclusion can be drawn from that.) Critical Uncertainties scenarios did not fare nearly so well; nor did econometric modelling.

In terms of the scale of the project. The largest projects, in terms of person-hours spent, were Europe 2000, OECD Interfutures, and Global 2000. The Polish project, The Year 2000, and Mankind 2000 were also large-scale operations. The smallest projects were Architecture 2000 and *As We May Think*, each involving a single person. Though one might expect that larger projects would produce closer matches, this was clearly not the case. If anything, the correlation was negative. A tentative reason for this is political pressure: a large-scale project, if centrally funded, is likely to face subtle pressures from its funders to produce acceptable findings, and to avoid unexpected outcomes. (There was evidence of this in the Norway 2000 study.)

In terms of the creators of the findings. The tentative conclusion from this very small sample is that the general public (in aggregate) is best at anticipation, famous people are worst, and futurists and experts are in the middle range.

AI.6 What other factors are related to foresightfulness?

This question covers two further points relating to the characteristics of a study: the time lag between anticipation and outcome, and the scope of the study. The following table presents the basic data for this analysis. The column headed Match has the same meaning as in the previous table, with a maximum of 5 symbols indicating the highest degree of matching, and one symbol indicating the lowest degree. The Scope column is a simple categorization: also ranging from 1 to 5 symbols, with the largest scope reserved for studies with global coverage geographically that also covered a very wide range of human activity. At the other extreme, a scope of 1 symbol (not in this sample) would apply to a single business in a single country. Telecom 2000 was scored at 2 points on this scale, because at the time of the study the sponsoring entity was a national monopoly: both one organization and one industry.

Time lag in the following table means the number of years between the year when the study was carried out, and the year for which the anticipation was made (generally 2000).

TABLE A1.4 FACTORS RELATED TO FORESIGHTFULNESS

Study	Scope	Time lag (years)	Match
Mankind 2000	****	33	*****
Architecture 2000	**	31	****
As We May Think	**	55	****
Australia at the Crossroads	***	20	****
Perspective 2000 (economic section)	***	12	***
Norway 2000	***	12	***
Poland 2000	***	13	***
Telecom 2000	**	25	**
Europe 2000	****	23	**
Kahn & Wiener	*****	33	**
OECD Interfutures	***	22	**
Internet commerce	***	3	**
Forecast 2000 (Gallup)	***	15	**
Global 2000	*****	22	*
Today Then	***	100	*

Relationship of scope to matching. Given the qualitative, subjective nature of the two scales, as well as the non-random nature of the sample, statistical testing of this relationship could have been misleading. However, inspection of the above table shows a slight negative relationship between scope and matching, muddled by a few exceptions, such as Mankind 2000. Other things being equal, it seems, a study with a very restricted scope tends to produce a closer match than a study conducted on a global scale.

In this case, there were two other plausible hypotheses that were not confirmed. The first was that foresightfulness would increase with scope, because perturbations at a small scale (such as for a single business) might be averaged out on a larger scale. The other unconfirmed possibility was for a curvilinear relationship, with middle-sized anticipations (such as for a region of a country, or an industry in a region) more successful than larger or smaller ones.

Relationship of time lag to matching. Considering the time lag between anticipation and outcome, two questions arose: (1) whether the later anticipations were more accurate than the earlier ones, and (2) whether there was any point at which the outcome in 2000 suddenly became much more obvious (though note, on the latter point that post-1995 anticipations, except in the area of ICT, were excluded for that very reason).

The scattergram below plots the Match variable against the time lag. The obvious hypothesis in this case is that the shorter the time-lag, the more accurate the anticipation. However, the above graph does not confirm the hypothesis at all, with a non-significant negative relationship between the two variables ($r = -0.14$) – though of course with such a tiny sample, the difference would have to be very marked indeed to be statistically significant. Removing the most distant outlier (the Galtung study, the only one with a Matching score of 5) makes the correlation slightly worse ($r = 0.21$) but still not significant at the 95% level of confidence. The conclusion is that there is no clear linear relationship between time lag and matching score.

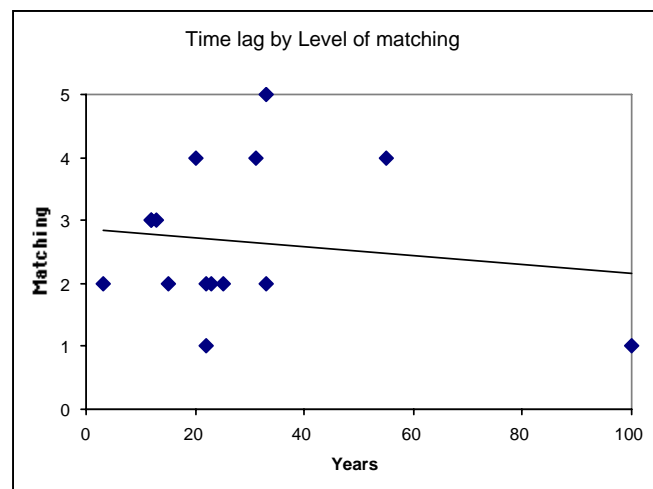


FIGURE A1.1 SCATTERGRAM OF MATCHING AGAINST TIME LAG

A1.7 What general problems of foresight were revealed in this review?

After working through the above review, various problems with the above anticipations became evident:

Problem 1: A restricted focus can obscure peripheral changes

The restricted focus of many of these scenarios prevented their creators from envisaging important peripheral influences – in one clear example, the fact that *Europe 2000* excluded Eastern Europe meant that the project was unable to foresee the possibility of a European Union enlarged to the east, with its attendant potentials and problems. The use of a method such as Critical System Heuristics (Ulrich, 1994), with its focus on establishing the boundaries of a project, could have helped in these cases.

2. Scenarios were more similar to their creation date than to 2000

These visions of the future predicted their own times better than they did 2000. Corn and Horrigan (1984) mentioned a study that foresaw high-speed overhead trains and airships reducing the journey time from New York to San Francisco to no more than 24 hours. Though it is not difficult to guess that this prediction was made *circa* 1930, it would be much harder in retrospect to guess which year the prediction applied to. Also, there were clear fashions in the focus of scenarios. In the 1960s, most were concerned with transport. In the 1970s, the focus moved to pollution and the environment, while in the 1980s it became computers. Scenarios mentioning globalization and sustainability belonged mostly to the 1990s. By paying so much attention to emerging trends and newly popular concepts, scenarioists risk overlooking longer-established factors.

Problem 3: Lack of justification for the target year

Though the scenarios were labelled as being for 2000, none of them supplied any intrinsic reason why they should apply in 2000 (rather than say 1990 or 2020). This problem is effectively an alternative to Problem 1: perhaps some of the errors described above were due not to poor foresight but to poor estimation of the target year. For example, *Australia at the Crossroads* (Kasper, Blandy, et al, 1980) predicted the world of 1985 more accurately than it predicted the world of 2000. However, a discussion with the second author revealed that the target year of 2000 was chosen almost arbitrarily, and there was no clear reason why the developments foreseen would take 20 years; in fact, many of them had occurred within five years.

Problem 4: Technological change exaggerated, but social change overlooked

A useful way to summarize the above scenarios is in terms of the error typology used in statistics, as described in most statistics textbooks (e.g. Abelson, 1995). When the true status of a null hypothesis (H_0) is compared with a decision, four possibilities exist, as shown in this contingency table:

TABLE A1.5 STANDARD CONTINGENCY TABLE

Decision	If H_0 is true	If H_0 is false
Fail to reject H_0	No error	Type II error
Reject H_0	Type I error	No error

Thus a type I error is a “false alarm” - the null hypothesis is true, but it is rejected, while a type II error is the failure to observe a true change. Applying this concept to the comparisons of scenarios with outcomes, a type I error corresponds to a foreseen situation that does not

occur, while a type II error is equivalent to the occurrence of an unforeseen situation. Thus those of Kahn and Wiener's (1967) 100 predictions for 2000 that did not occur - such as human hibernation - were type I errors, while the failure of *Poland 2000* to foresee the end of communism was a type II error.

Applying this typology to the above studies, it is evident that (in general) technological change was overestimated (type I errors) while social change was overlooked (type II errors). It is useful to consider why this should be so. The over-prediction of technological change seems to be related to the social climate at the time and place of the scenarios' creation. For example, Kahn and Wiener's confident predictions about technology were not made in isolation. Particularly in mid-20th-century USA, there was wide discussion and public awareness of new technological developments. During that period, technological solutions were widely expected; for example, there was much focus on speed records. Ascher (1981) notes the extremely high forecasts for general aviation made in the 1950s for the 1960s: at that time, it was expected in North America that many men would commute by air, often piloting their private helicopters.

Against this demand factor (public expectation of rapid technological development) was a wide range of supply-related factors:

1. It takes many years to develop a technology from a demonstration level (such as one moon-landing) to wide use (mass holidays on the moon);
2. For communications-related technologies, Metcalfe's law (Metcalfe, 1995) applies: that the value of a communications system is proportional to the square of the number of participants. Thus fax machines took many decades to become widely used outside the newspaper industry; there is no value in owning the only fax machine in an inter-communicating group (Petroski, 1996, pp104-119).
3. A chicken-and-egg law applies to sets of technologies that are inter-related. For example, compact disc players are useless without compact discs, and without the efforts of Philips Electrical and Sony – active in both the electronics and music industries – the quick success of this medium may not have occurred. (The other prerequisites, such as audio components, a stable electricity supply, a retail distribution system, and public interest in owning recorded music by that already existed.)
4. Even if a technology exists, the price may be too high for would-be users, and only the eventual interaction of demand with economies of scale can establish financial viability.

5. Governments, influenced by vested interests, play a strong role in technology adoption, even if strong demand exists. Examples are the long-delayed introduction of television in South Africa under the *apartheid* government, and the legislated ending of analog television broadcasts in various countries around 2010.

6. Even if all the above factors are favourable, there may be insufficient demand for a technology: for example, the video telephone (Schnaars and Wymbs, 2004).

All six factors can cause delays, but there exist few counterbalancing factors that could accelerate the adoption of technologies. One such factor is a national sense of urgency in response to threat, involving government control and massive funding, exemplified by the NASA moon-shot in 1969, the result of a ten-year plan in response to the USSR's satellite launchings in the late 1950s.

Turning next to the reasons for the under-prediction of social change, let us return to the example of commuting by helicopter. One illustration to a future-related article from a US publication *circa* 1960 (which I have been unable to find again, to cite) shows a man taking off in his private helicopter from the front lawn of his house, leaving for work. His wife, wearing an apron, is waving goodbye. In this vision for 2000, men commute by helicopter, while women stay home to do the housework: a classic example of technological over-prediction and social under-prediction. It is useful to reflect on why this error was made. Because the author's focus was solely on technology, the possibility that by 2000 most women might be working was completely overlooked. However, even in terms of the mindset of the 1950s in the USA, the idea that most women would be working by 2000 was probably quite conceivable. What may *not* have been conceivable in the 1950s was an illustration of the managerial woman taking off to work in her own helicopter, while the "help" (probably a Hispanic woman: in the enlightened twenty-first century, there is no "servant problem") waved goodbye. Factors causing the growing disparity of incomes that would enable such an illustration may not have been visible in the 1950s.

Problem 5: Scenarios were often not distinct from each other

From the 1980s onward, most of the references were ensembles of scenarios (rather than single predictions). Often, all the scenarios in an ensemble turned out to be fulfilled to some extent – and, from a post-2000 perspective, all the scenarios in an ensemble often seemed very similar to one another, as noted by Liebl (2002).

Problem 6: Outcomes were specified too vaguely to evaluate

Some scenarios and predictions were written so imprecisely that it was impossible to determine whether they had been achieved or not. Often this arose through a failure to quantify. For example, to say “the level of public health will be higher in 2000” is all but meaningless, because the key variable was not described fully and not even approximate quantification was attempted. Unless some indication is given of prevalence, such statements are almost meaningless.

Problem 7: End-state scenarios were difficult to envisage, lacking detail at human scale

A scenario is intended to be a description of a unified world, but on reading the end-state scenarios, it became obvious that lack of provenance led to a lack of clarity. Thus some scenarios were only vague descriptions of ill-defined attributes; an example is some of the variant scenarios described by Kahn and Wiener (1967). The few chain (sequential) scenarios presented clearer pictures of possible worlds, even though the descriptions of successive states in the chain were usually sketchy. However unlikely these might have seemed in retrospect (e.g. some of Herman Kahn’s scenarios for World War III), and however unpopular, they at least had enough coherence and plausibility to provide a basis for action.

Problem 8: Overlooking the personal perspective

Looking at the domains on which the scenarios focused, a concentration on external issues and the macro scale was evident: these were scenarios of the etic rather than the emic, *Gesellschaft* rather than *Gemeinschaft*. Of the studies in Table 2, only that of Jungk and Galtung (1969) extended its coverage to the ways in which people would live and think in 2000. If one accepts the view taken by thinkers including St Augustine (397), G H Mead (1932), and Jaques (1982), that the future exists only in people’s minds, to omit any consideration of those minds in 2000 seems a serious omission - particularly if it is also accepted that expectations can create the past as well as the future (Michael, 1985).

Problems noted by others

Some related problems discussed in other writings on the comparison between anticipations and outcomes have been discussed by other writers. Galtung, in his review of expert predictions for 2000, found that “the basic mistake made by the experts quoted is their inability to identify countertrends” (Galtung 2003b:129n). In the present study, such countertrends were not clearly evident, except perhaps the relative decline of the Japanese economy following its

“bubble” in the late 1980s. Several of the above studies, such as Kahn and Wiener (1967) and Interfutures predicted a much stronger economic ranking for Japan in 2000 than eventually was the case.

Kahn and Wiener (1967) include a section entitled “Ways to go wrong.” They list ten of these: criteria too narrow; decisions at inappropriate point in the structure; inadequate thought; bad luck; unknown issues or unlikely events; changes in actors; inappropriate models; inappropriate values; over- or under-discounting of uncertainty or the future; and “the best may be the enemy of the good” (implications of which were not clearly defined).

Cornish (1997, p48) identifies as the main problem “shifts in the factors underlying the trends that [the forecasters] had projected into the future.” For example, the drop in space exploration funding from the 1970s meant that predictions of human colonies on the moon by the 1980s were not borne out. Another common problem identified by Cornish was predicting a trend correctly, but overestimating its rate of diffusion. This is another aspect of problem 3 above.

AI.8 What implications for foresighting arise from this study?

For each of the eight problems listed above, this section suggests some possible solutions.

Problem 1. A restricted focus can obscure peripheral changes

A solution is the use of methods such as Critical System Heuristics (Ulrich, 1994), and the boundary critique of Midgley (2000). These have been successful in establishing the boundaries of projects in the “messy” areas of human services. For those studies that had problems with boundaries, the problem was always that boundaries were too narrow, so a simple solution might be to begin a study with boundaries a little broader than initially seem necessary. A related solution used by Ulrich and Midgley is to begin a study by questioning its boundaries. This might usefully be done by reviewing first-order and second-order stakeholders.

Problem 2. Scenarios were more similar to their creation date than to 2000

Underlying this problem may have been that some aspects of issues studied seemed unimportant to the creators of the scenarios, and they thus assumed there would be no change on those variables - as in the case of the male helicopter commuter.

Solution 1: Focus on sequences of anticipated change, using a multiple perspectives approach akin to that of Linstone (1984), but applying it to multiple perspectives of times – looking both forward and back in time, both rooting the future in the past (as did Kahn and Wiener, 1967), and focusing on sources of potential discontinuity rather than continuing trends.

Solution 2: As for problem 1: avoid being restricted to too narrow a scope. Though there is clearly a danger in broadening the scope of a study too much, many of the problems found in these 15 studies resulted from terms of reference that were too restrictive.

Problem 3. Lack of justification for the target year

At least in the area of technology, it has been more difficult to predict when something will happen than what will happen. (To continue an earlier example: interactive television may yet succeed.) Technology roadmapping (Phaal et.al, 2004) offers a short-term solution, for time-scales of only a few years. However perhaps a precise target year is often unnecessary, given that a frequent purpose of scenarios is to affect decisions in the near future, well before the target year is reached. For example, *Australia at the Crossroads* was adopted up by a recently elected government soon after 1983, well before its 2000 target date. A simple solution is that, instead of a target date, a range of dates, or an earliest-possible date could be estimated.

Problem 4. Technological change exaggerated, but social change overlooked

Each component of this double problem can be addressed separately. Firstly, a possible solution to exaggerating technological change is to take into account delays caused by technical coordination and market take-up rates. Methods that may be helpful are backcasting, technology roadmapping, and historical comparisons using concepts such as the empirically well-supported Bass curve for the diffusion of innovations (Mahajan, Muller, and Bass, 1990).

The problem of overlooking social change appears to be related to the focus and scope of studies: participants do not notice what they were not looking for. A possible solution is thus the same as that for Problem 1: willingness to enlarge the scope of a study.

Problem 5. Scenarios were often not distinct from each other

This seemed to occur particularly when the Critical Uncertainties method was used, and scenarios were based on extremes on a two-dimensional quadrant. This might be overcome by using a morphological relevance-tree approach to scenario construction, rather than a dimensional one. Relevance-tree classification, being hierarchical is inherently mutually exclusive.

However this cannot be a complete solution, because there remain the problems of the sequence in which variables are split, and the method of splitting when a variable is not Boolean.

Problem 6. Outcomes were specified too vaguely to evaluate.

One alternative to making a “safe” statement such as “the level of public health will be higher in 2000” is to specify expected incidence levels for particular health conditions. However, this faces the danger seen in the Telecom Australia study: precise prediction of a variable that will no longer be relevant. A preferred approach may therefore be to anticipate variables rather than values: in this example, and/or to list the conditions expected to be reduced below specified threshold levels in the target year.

Problem 7. End-state scenarios were difficult to envisage, lacking detail at human scale.

To increase vividness, in two aspects: both time-based (how exactly the present could lead to an envisaged future) and context-based (how that scenario would fit in with other aspects of life at the target time). Incasting, as used by Dator and Schultz (Schultz, 2003) can be helpful in this regard.

Problem 8. Overlooking the lifeworld perspective.

Perhaps because many of these studies were carried out by economists rather than psychologists, all but one (that of Ornauer and Galtung) focused on social rather than individual futures. However, some recent writers on foresight and related issues (Slaughter, 1995a, 2004; Wilber, 1997, Voros, 2001) have noticed the omissions and have tried to generate more person-focused futures. Again, methods such as incasting (Schultz, 2003) are helpful in this regard, as is the inclusion of citizen participation in foresight work.

Conclusions

Eight problems were found in this study, with tentative solutions as suggested above. In addition, three further observations were made, in the form of lessons learned.

Lesson 1: Intuitive scenario methods produce closer matching than critical uncertainties and econometric methods.

In scenario planning, the Critical Uncertainties method did not perform well (at least, on this outcome-based criterion - regardless of what it might achieve in terms of immediate impact

on client organizations). Intuitive approaches to scenario building were more effective. Their value is widely accepted (e.g. van der Heijden, 1996; Schwartz, 1991) - with the proviso that success with these requires an unusually high degree of skill and futures awareness.

Lesson 2: Large-scale scenario exercises (involving more than 5 person-years' work) were no more effective than smaller-scale exercises.

As there was no correlation between the scale of a project and its accuracy, if the latter is to be the key criterion there appears to be little point in organizing large-scale foresighting exercises. The same budget might be better spent on a variety of small, independent studies – provided that these were not subject to political pressure to develop futures to please the clients.

Lesson 3: Include a wide range of participants, but avoid famous people with vested interests and experts in narrow subject areas.

Since most of the studies that used input from experts did not fare well at anticipating 2000, there is clear danger in relying too much on experts. The problem appears to be that experts, by definition, have a narrow focus: they are experts *on* something. When the background to that situation changes, their expertise may no longer be relevant. This is borne out by some of the well-known quotations cited in Cerf and Navasky (1984) and Shuman (1997), including...

- “Aircraft are interesting toys, but of no military value.” - Marshal Foch, France, 1912.
- “I think there is a world market for as many as five computers.” - Thomas Watson, head of IBM, 1943.
- “There is no reason anyone would want a computer in their home.” - Ken Olsen, founder of Digital Equipment Corporation, 1977.

These people were among the foremost experts of their time. Foch was one of the most progressive generals of the French army, Watson was the founder of IBM, and Olsen the founder of the Digital Equipment Corporation. All of them were proved wrong within a handful of years, yet the prognoses of Vannevar Bush, a more generalist expert, are still foresightful even now.

To summarize the above three lessons, it follows that an effective method of foresighting would use intuitive methods (though perhaps not solely), it would avoid using famous people with vested interests and experts in narrow fields, and if sufficient funds were available they would be used for several independent small studies rather than one large one. These conclusions, together with the tentative solutions to the problems found, are applied in chapter 3.