Martin Bauer and John Durant

Mapping European public understanding of science: a re-analysis of the open question included in the eurobarometer survey no 31 on S&T from 1989

Report

DG 12 of the Commission of the European Communities

Mapping European Public Understanding of Science

A Re-Analysis of the open question included in the

Eurobarometer survey no 31 on S&T from 1989.

Martin Bauer
John Durant

in collaboration with

A Allansdottir, P Fluegel, M C Gervais, S Jovchelovitch, A M Mello
A Porter, I Schoon, A Stathopoulou

The Science Museum
Exhibition Road
London SW7 2DD

Tel 071 938 8237
Fax 071 938 8213

9 July 1992

Contract No PSS*0434

Content

Summary

1. The problem

2. Eurobarometer no 31 from March/April 1989

   2.1 The survey data
   2.2 The translation of the open question
   2.3 The coding of the open question

3. The Re-Analysis

   3.1 The coders
   3.2 Reliability
   3.3 Problems

4. Results

   4.1 Technical problems in re-integrating the raw data
   4.2 American approach: 'attentive public' and 'science literacy'
   4.3 Problems with ranking the nations on PUS
   4.4 Towards a map of European Public Understanding of Science
      4.4.1 The approach of the Science Museum London
      4.4.2 The Science Museum Coding Frame
      4.4.3 The Reliability of Science Museum Coding
      4.4.4 The Results of Science Museum Coding

   4.5 Comparing US and Europe

5. Recommendations

6. Literature

7. Appendix

   7.1 The coding frame as reported by 'Faits et Opinions' 1989
   7.2 The coding frame as reported by Gallup GB
   7.3 The coding frame used for the US88, US90 and GB 1988
   7.4 The Science Museum Coding Frame
   7.5 The Items of the Knowledge Scale
   7.6 Selected data tables
   7.7 The frequency distributions of the variables

Note

Summary

The present investigation starts from serious doubts about the reliability of one particular element within the Eurobarometer survey no 31. This survey had among other purposes given the basis for a comparison of European and American Public Understanding of Science. The focus of the analysis is the translation, transcription and coding of the open question, "what does it mean to study something scientifically?".

Errors in the translation and the non-standardized procedures for the transcription and the coding of verbatim responses are matters for special attention in future surveys. The problems do not disqualify the open question per se as a useful source of information. Rather the way the question is handled needs improvement. Procedures have to be more strictly standardized across countries, and a different coding frame is proposed.

The ranking of nations on PUS measures proves statistically futile. First, the ranking, particularly in the top group, depends strongly on the kind of scale used. Second, the differences of average scores are not significant. The ranking along observed differences is therefore statistical nonsense. A grouping of nations is a more accurate way of presenting results.

The American approach to measuring the 'attentive public' and 'science literacy' is problematic, because neither concept is unambiguously defined at the level of data (see chapter 4.2). For valid international comparisons a more rigorous approach is needed.

The approach of the PUS research group at the Science Museum, London, introduces the concept of Public Understanding of Science as a three dimensional concept: factual knowledge (facts), methodological knowledge (process) and knowledge about how science is organized (institution). Each dimension must be measured with a reliable scale. Rather than comparing mean values across nations, it is more fruitful to look at structures between variables and their variation across nations. Particularly the bi-modality and the variance of the knowledge distribution is important to characterize the science culture of a nation. It is more accurate to cluster nations with the help of several variables than to rank them on insignificantly different average scores.

International comparison of data must be able to map qualitative differences among cultures, particularly in a multi-lingual context like the EC. The careful analysis of the open question is a step in that right direction. The Science Museum Coding Frame for the open question accomplishes to map qualitative differences of answers on four dimensions: process, institution, effect, and examples. Correspondence analysis of these variables reveals a coherent pattern with regard to how strong a Popperian notion of science people have (process), what kind of institution they think 'science' is (institution), what kind of effect people attribute to science and whether that effect is positive or negative (effect), and what examples are paradigmatic for science (example). The US and Britain are similar in their outlook on science. Continental Europe is different in many respects. This historical fact is clearly represented in our

data.

Comparing the US and Europe one can say that the American public is on average more attentive to science and technology than the Europeans. The 'popperian' view of science, stressing theory, hypothesis testing, experimentation and measurement, is more wide spread in the USA than in Europe. The Europeans give more complicated answers to the question 'what does it mean to study something scientifically', provided they give an answer at all. Generally difference among the EC countries are greater than between the EC and the US. We end the report with 12 recommendations for future surveys on Public Understanding of Science.

'If you understand something in only one way,
then you scarcely understand it at all'
(M Minsky, 1992)

1. The Problem

In the course of analyzing the Eurobarometer no 31 data (Bauer et al. 1991; Durant et al., 1991) on European Public Understanding of Science and Technology several problems have arisen which cast serious doubts on the reliability, of the open question q452/453, '**please tell me in your own words, what does it mean to 'study something scientifically**' ?, for international comparisons:

- The expected correlation between process measures and knowledge measures did not materialize in the European data (GB 1988: correlation r = .68; EC 1989 correlation r = .49)
- The coding frame used for the 1989 data does not correspond to the US frame as used by Miller. Different agencies have used different coding procedures.
- The coding process as is was coordinated by Faits et Opinions, Paris, is not transparent and insufficiently documented. It is unclear, a) how different agencies coded the open question, and b) how the different coding procedures were integrated into the one variable of the Eurobarometer data set.
- The language and cultural variety in Europe makes it problematic to impose an anglo-saxon meaning of science as the baseline to assess the level of scientific understanding of science in other countries. An approach that is more sensitive to semantical differences seems preferable in the European context.

We took these problems not as an indicator of the failure to assess public understanding of science, but as an opportunity to evaluate and to improve the measures used. The open question, which has been in use since 1959 (Withey, 1959), with nearly 10000 responses in 9 different languages, is a very rich source of data. In February 1991 we approached the EC DG12 to conduct this project to reanalyse the verbatim responses of the open question of the Eurobarometer no 31 from spring 1989.

Vocabulary

In order to be clear throughout the report we set the following conventions for technical language:

Open question: All questions of the Eurobarometer no 31 S&T section were closed questions, with one exception. This single open question is the main object of the present analysis

Items: 'Item' is the technical term for survey questions that are combined to a scale or a measurement dimension.

Verbatim response: All answers to the open question have been recorded in one way or another. This record is referred to as a verbatim response.

Coding frame: The guideline which coders use to turn verbal responses into numerical data for statistical analysis. A coding frame reduces complexity for a certain purpose.

US coding frame: The coding frame that has been introduced by Jon Miller for the analysis of the open survey question.

SM coding frame: The alternative coding frame that has been developed by the team at the Science Museum (SM), London, for the same open survey question.

Reliability: The quality index of a coding frame. It indicates that the same coder produce roughly the same measures in the same context. We are using Kappa, that measures the strength of association between two categorical variables taking into account random allocation.

Raw data: All verbatim responses are coded on a coding frame. Raw data refers to the numerically coded answers to the open question.

Facts: The core dimension of the concept of Public Understanding of Science consists of a number of questions on factual knowledge.

Process: The second dimension of the concept of Public Understanding of Science consists of a number of questions on scientific methods.

Old data: The original data file of Eurobarometer no 31, that is publicly available from the data archive ZUMA in Mannheim, Germany.

New data: The data that is generated by the recoding of the responses to the open question done in this project.

Data integration: The combination of the old data and the new data that is created by this project.

Correlation: A measure of covariance between two variables. If one variable has a high value the other is likely to have a high value as well. This relationship is expressed by an index between -1 and 1. The strength of that relationship varies. -1 or 1 means a perfect relationship. 0 means no relationship at all. We are using here Pearson's Product-Moment-Correlation index. Other measures are rank-correlations, e.g. Spearman's, which measure the matching of two rank orders with an index of maximum value 1.

2. Eurobarometer no 31 from March/April 1989

Eurobarometer no 31 was conducted in the 12 EC member states in March/April 1989. In each country a 1000 sample of the population above 15 years of age was taken. Exceptions are Luxembourg with n=300 and Northern Ireland with n=300, which is a sample taken separately from GB. The total sample size is n=11678. The sample was regionally stratified in about 1350 sampling points, with a mixture of quota and probability sampling, depending on the infrastructure of the countries (Eurobarometer no31, 1989, A4).

The international coordination of the survey was done by **Faits et Opinions, Paris**. Eurobarometer has since been relaunched by a different company INRA, Brussels, which brought a number of changes to the way the survey is conducted.

2.1 The data

The structure of the survey follows the model that had been used before in the US and in the UK (Miller, 1983; Durant et al., 1989; Evens et al. 1989). The core consists of a number of items on factual knowledge (the Oxford Scientific Knowledge Quiz) and items on scientific methods. These two dimensions operationalize **the concept of Public Understanding of Science (PUS)**. We recommended in a report to the National Science Foundation, Washington, to add a **third dimension** in future surveys: knowledge about the **organisation of scientific research** (Bauer and Durant, 1991). The key concept has proven powerful in predicting interest in and attitudes to various areas of science and technology, as well as confidence in European science in relation to Japan and the US. The open question, which mainly concerns us here, is an item of the process dimension of PUS. The **Science and Technology section** of Eurobarometer no 31 includes the variables v14-v41; v138-v200; and v377-v399[1]. The open question comprises the variables v394-v399 of the old data file.

The **uncoded verbatim responses**, as archived at DG 12 in Brussels, are incomplete. Three data sets are missing: Great Britain, Belgium and Denmark. The British data were destroyed in December 1989 as no other instruction had been received by Gallup. The Belgian and Danish data cannot be traced either. The **status of the verbatim data** varies: loose single pages of the completed questionnaire (Italy, Northern Ireland, Ireland); cut off slips of the question (Greece); handwritten transcripts on separate sheets of papers (France, Germany); computer printout (Luxembourg); typed and bound transcripts in report form (Spain, Portugal, Netherlands). It is recommended that a standard format be adopted for the presentation of the verbatim responses in the future - typed and bound transcripts - together with clear instructions on how to archive the responses. For comparison we incorporate the American data from the surveys of 1988 (US88) and 1990 (US90) in our analysis. The American data

[1] The survey includes questions q130-q160; q235-271; q278-280; and q443-q453 of the original English and French questionnaire

is available as a computer printout of verbatim transcriptions of the answers[2].

[2] We should like to thank Jon Miller and his colleagues for the help that they provided in accessing that data.

2.2 The translation of the open question

The translation of questions is a major difficulty of multi-national survey research. Only an equivalent translation of a question yields comparable results. If the translation is inaccurate we measure semantic differences instead of differences in opinion, attitudes and behaviour.

Table 1: translations of the open question

Original form: conceptual version

GB and Ireland: Please tell me, **in your own words**, what does it mean to 'study something scientifically' ?

Lux: Kennt Dir mat **Eren eegene Wierder** soen wat 'eppes wessenschaftlech studeiren' heescht ?

Port: Pode dizer-me **por palavras suas** o que significa 'estudar qualquer coisa cientificamente' ?

E: Digame por favor, con **sus propias palabras** que significa 'estudiar algo cientificamente' ?

G: Bitte sagen Sie mir in **Ihren eigenen Worten**, was es heisst, 'etwas wissenschaftlich zu untersuchen' ?

Personalized form: experiential version

F: Pouvez-vous me dire **dans vos propres termes** ce que cela signifie **pour vous** 'd'etudier quelque chose scientifiquement' ?

I: Puo dirmi con **le sue parole** cio che significa **per Lei** 'studiare qualcosa scientificamente' ?

Greece: Could you please tell me **in your own words**, what does it mean **if you study** something scientifically ? (translated)

NL: Kunt u mij in **uw eigen woorden** vertellen wat **voor u** de betekenis is van 'iets wetenschappelijk bestuderen' ?

The translation procedure of the Eurobarometer no 31 is not documented (Eurobarometer no 31, 1989). Examining the translations of the question in the 9 European languages, as shown in table 1, reveals an error which may have introduced a bias into the results. The translations differ in the degree of **personalization of the question**. The original question, correctly translated in German, Spanish, Portugese, and Luxembourg, asked respondents to express in their own words the meaning of 'studying something scientifically'. The respondent gives the personal formulation of the **general concept** 'scientific method'. In the Dutch, French, Greek and Italian case we have a stronger personalization of that question: 'in your own words' and 'for you'. The 'for you' links the question to personal experience. Hence, the respondents are asked to formulate their **personal experience**. This **double personalization** of the question makes a semantic difference that most likely affects the responses. Personal experience is normally easier to formulate than a general concept. Our coders have observed more answers of the 'I would do ...' type in the second case.

We tested the **translation effect** on the American coding frame by grouping our results according to table 1. The formulation of the question was conceptual for US, GB, Ireland, Northern Ireland, Luxembourg, Portugal, Spain and Germany. The question was experiential for France, Italy, Greece, and Netherland. The average coding for 'conceptual' is 3.59 (n=5900), the average coding 'experimental' is 3.48 (n=2905)[3]. The difference is significant (Oneway analysis of variance: F=16.9; p=.0000; df=8804[4]). The incorrectly translated version of the question produces better results than the original version of the question. Hence we have good reason to suspect a translation effect at work that introduces a bias to the results of some countries. The comparability of the open question from Eurobarometer no 31 across all EC countries is problematic[5]. This does not disqualify the viability of the open question in general. The translation needs to be accurate in the future. That problem of question wording may be tested in a future survey by a split half design with two versions of the question.

2.3 The coding process

Coding reduces the complexity of the verbatim responses for a specified purpose. Several problems arise. We lack any documentation on how these problems have been solved in the coding of the open question for the 1989 data. In order to compare the data across countries standardized procedures are necessary. We suspect that neither the transcription nor the coding of the respondents' answers were sufficiently

[3] The lower the score the better is the methodological knowledge according the American coding frame: 1 = high; 5 = low understanding (see appendix b)

[4] Significance tests are normally documented with three parameters: the value of the test statistic (F or t), the probability of that value, and the degree of freedom, which related to the number of observations.

[5] It is, however, not decidable whether the translation effect is the only factor that explains the difference, because other factor could not be controlled in the analysis. In order to avoid this possible confounding effect, more care needs to be taken in translating the question correctly.

standardized.

The transcript of the response

In the Eurobarometer the interviewer confronts the respondent face-to-face. The US data is collected by telephone interviews. The interviewer records the answers of the respondent. How does he or she do that ?

First, is the transcription verbatim, or is it a summary ? In the US procedure a summary of the responses is made. Interviewers receive training in summarizing accurately open answers. The answer is typed into the computer by the interviewer. The space allocated for the record is 80 characters. Later the coding is done by another person[6]. The records are summaries or partial transcriptions of the responses, many responses are incompletely recorded. The length of the transcripts is limited to 80 characters. The truncation of American verbatim responses for technical or practical reasons may not give the full richness of many of people's responses. We are aware of that possible limitation. However, on examining the US data, we assume that the overall results would not be very different, if responses were complete. For the future fielding of the open question it is advisable for purposes of comparison to avoid that limitation in the data collection. We have no record about the European practice. Enquiries after three years only yield a formulation of the companies' general policy. We conclude from the inspection of the European raw material that it was handled differently by different subcontractors. For some countries whole phrases are recorded, for others only short propositions or single words. The length of the answers generally varies between countries and not only between respondents. This indicates the varied transcription procedures.

Second, are the interviewer and the coder the same person ? If the interviewer is aware of the coding frame, the summaries of the responses will be structured by the knowledge of the coding frame. Again the European practice is not documented and likely to be not standardized. Short and stereotypical records of in the Irish, Italian and Portugese data indicate that the summary was made by the coder. Recording and coding seem to derive from different persons in the other cases.

The coding of the verbatim responses

Once the responses are recorded, they have to be coded to reduce the complexity and to construct numerical data. Two questions arise: a) Is the coding frame well defined and transparent ?; and b) is the coding process reliable? The coding frame used for the Eurobarometer no 31 is supposed to match the American frame used by Jon Miller. Scholarly critique about the lack of technical reports and about ambiguous definitions of measures have improved the transparency of the American data (Beveridge and Rudell, 1988). The transcription and the coding of the answers is done by different persons. The **procedure seem to have changes since 1983**. Miller reports in 1973 'responses were later coded by two coders independently, and those cases that involved disagreements in coding were judged by the coding supervisor' (Miller, 1983, 37). Our enquiries have shown that J Miller is doing the coding of the open answers

[6] This was confirmed in a telephone inquire and with a fax from the Public Opinion Laboratory, Northern Illinois University, DeKalb, on 7-Jan-1992.

personally in recent surveys to guarantee expert ratings. Despite the fact that two coders handled the material in 1983 the level of reliability was not assessed. An reliability index evaluates the viability of a coding frame for other contexts and is necessary for the standardization of the method.

The practice used to code the European data from 1989 is unknown. No report from Faits et Opinions describes the treatment of the open question. The coding frame used for the UK and the coding frame used in other countries are incompatible. It is impossible to see how the one can be transformed into the other (see appendix for the two different coding frames). Nevertheless it has been done. The resulting noise explains the low correlation with other measures of PUS. People from Faits et Opinions were aware of the problems with the open question and have informally urged caution when analyzing that data. The reliability of the open question of Eurobarometer no 31 is therefore equally unknown. We regard the Eurobarometer data of variables **v394-399 as being unsuitable for any analysis**. Again this is not a verdict on the viability of the open question, rather on the way the data collection was conducted.

We conclude that the open question, as it was recorded and coded in 1989, is unreliable and should not be used for any analysis. We recommend to keep the open question as an important source of first hand semantic information on Public Understanding of Science and to introduce **transparency and standardization** in the transcription and coding procedures of the open question in the following order of priority:

- A standard and typed record of the verbatim responses is needed
- The coding frame and definitions need to be documented
- Interviewer and coder should different people, if not that should be mentioned
- The interviewer must not be aware of the coding frame. If he or she is aware of it, that should be mentioned.
- The coding procedure needs documentation
- The Reliability of the coding process must be calculated

3. The Re-Analysis

3.1 The coders and the coding procedure

For the re-analysis of the open question a multi-lingual team of coders was formed and moderated by the authors. Each coder was responsible for one of the nine languages: Greek, Spanish, Portugese, German, Italian, Dutch, English, Luxembourgish, French[7]. The team members were all PhD students and were, with two exceptions, from the London School of Economics, Department of Social Psychology. All had experience with content analysis of various materials. The team met several times to familiarize themselves with the verbatim material and the aims of the project. We discussed the **American coding frame** (see appendix 7.3). Each coder took the material away for coding. We met again several times to discuss the problems and our observations. Coders were asked to note any observation that they made during the coding process. With the brain storming method and the guidelines of some theoretical ideas we developed The Science Museum Coding Frame for the open question and recoded the material. We assessed the reliability both of the American and of the **Science Museum Coding Frame** (see appendix 7.4).

[7] We should like to express our thanks to Agnes Allansdottir (Italian), Peter Fluegel (German), Marie-Claude Gervais (French), Sandra Jovchelovitch (Portugese), Ana-Maria Mello (Spanish), Alan Porter (English and American), Ingrid Schoon (Dutch and Luxembourgish), and Angela Stathopoulou (Greek) for their cooperation. In an unprecedented spirit of European Community the team has made important contributions to the critique of the American coding frame and the development of the alternative European coding frame.

3.2 Reliability of the American coding frame

Reliability is a quality criterion of a measurement instrument. An acceptable reliability demonstrates that the measure is consistent when obtained repeatedly, in different contexts and by different coders. Only a reliable measure is a useful measure. In the present study the reliability was measured without intensive coder training. Coding frames for surveys must be simple and immediately applicable by coders. A sample of the Irish, German and Dutch data (n=378) was recoded by another coder a month later.

The **inter-rater reliability** was measured[8] with kappa (Cohen, 1960) and with Spearman's correlation coefficient as shown in table 2. Kappa measures the coder agreement between the first and the second coding against the assumption of random coding. The American coding frame produces an ordinal scale of the respondent's graded knowledge of the scientific method. Spearman's r measures the rank correlation of both codings. The reliability is similar across different countries and different coders. The total reliability of the coding frame is however rather weak with kappa=.44 and Spearman's r=.62.

Table 2: Reliability of the American Coding Frame

variable	kappa	Spearman's	n
Total	k = .44	r = .62	378
Irish	k = .48	r = .65	144
German	k = .48	r = .69	104
Dutch	k = .37	r = .63	128

Krippendorf (1980, 147) defines an 'acceptable' level of reliability for cultural indicators as .80. Reliability in the .60s is acceptable for tentative conclusions, when the implications are not very costly. We regard the level of reliability of the American coding frame as **doubtful** for our purposes. Again this by no means disqualifies the viability of the open question. Rather this indicates that a) the categories of the American coding frame are not clearly defined or b) the categories of the American coding frame does not do justice to the responses given. We recommend to clarify the definitions of the US coding frame as well as to pursue alternative ways of analysis, particularly in the European multilingual context.

[8] The software package SPSS was used for the calculation of kappa and Spearmans r. Kappa assumes categorical data, Spearman's r assumes rank order data.

3.3 Problems with the secondary analysis of the data

A number of problems surfaced during the coding process. First, the American coding frame seems inadequate for the responses. According to the report of our 8 coders most answers have to be 'squeezed' into the categories. Albeit, it seems less difficult for the British, Northern Irish, and USA data. This points to a problem of language and culture. The American coding frame has an Anglo-Saxon bias with regard to the meaning of scientific study. The coding frame implicitly endorses a normative philosophy of science that one could briefly characterize as 'Popperian' (e.g. Popper, 1963): a notion of science which focuses on theory building, hypothesis testing, measurement and experimentation. Science as a normatively defined canon of methodology and procedures is both historically and sociologically unrealistic, and seems not to be in people's minds either (Harre, 1972).

If one wants to go beyond measuring the diffusion of a single notion of scientific method, this has implications for the construction of the coding frame. The problems of applying the American coding frame indicates the inadequacy of squeezing different notions of science, more or less elaborated in the responses given, into the 'Procrustes bed' of a single notion.

Second, people **generally have difficulty in answering the question in a grammatically correct way**. One would expect respondents to give a verbal answer plus an adverbial qualification to the open question. The answers would have the form: 'to do X in a Y way'. Most of the answers do not follow this expectation, rather they have the character of **free associations.** Two examples illustrate this point: an answer of a Greek woman - 'When my husband lived, who was a lawyer, he used to explain everything to me'; and an Irish one - 'to get a white coat .. to get a job with ICI'. These responses are significant in their own right for the public understanding of science. They show a Greek woman's world view that science is a thing for knowledgable husbands or an issue of social positioning in the Irish case. Such answers are usually coded with the rest category 'other'. The American coding frame, useful as it may be, is not complex enough to make sense of such responses. The information that resides in the answers is lost.

Third, the rest category 'other' contains answers about **Public Representations of Science** (Durant et al., 1992) which are masked as mere noise. As a matter of fact the rest category contains other dimensions of understanding, not only nothing of the one dimension.

Fourth, **the sophistication of the answer is not reflected in the coding frame**. The coding accorded to key words like 'theory', 'experiment', 'test' may lead to codings of doubtful validity. Simple and sophisticated answers are coded in the same category because of the mentioning of a certain word. A sophisticated answer may be coded low, because a key word does not appear.

Fifth, the location of the open question at the end of the survey may influence the responses (**context effect)**. Coders have reported that examples from metallurgy are frequent. It is unlikely that such a special area of research like metallurgy is salient in public minds. This rather reflects the design of the survey. A split half design was used to test the understanding of experimental method with a medical and a metallurgical example.

All these observation suggest the need for a revision of the coding frame a) to accommodate the different 'natural' meanings of 'scientific method' in different languages, and b) to catch the sophistication of responses more faithfully to the data. Too much reduction of complexity does not do justice to the international and multi-cultural public.

4. Results

4.1 Technical problems in re-integrating the raw data

We encountered two problems in re-integrating the recoded new data with the old data file. One could be pragmatically solved, the other could not be resolved.

First, the **verbatim responses were incomplete**. The British, the Danish and the Belgian responses had been lost. This has implications for the comparison of aggregates between Europe and the US. From previous analysis we know that GB and Denmark are above average in public understanding, while Belgium is below average (Bauer et.al, 1991)k, and we also know about the correlation between factual knowledge and methodological knowledge (Durant et.al, 1989). When two above average countries against one below average are missing it is inadequate to compare the European data with US data. The European score would be underestimated. The following procedure was used to compensate for the missing data.

The missing British data was substituted by the data from the GB survey of 1988 (Durant et al. 1989). We assume that the measures on PUS are fairly stable, and changes are unlikely to occur within the period of one year on the aggregate level. Hence, we neglect the one year difference between the two surveys. The British data had a double size sample of 2000 instead of 1000. Knowing about the correlation between factual and methodological knowledge on science, we hypothesize that this compensates for the missing Danish data in the European aggregate, because both countries are above average in factual knowledge of science (see appendix table 8).

The British data of 1988 (GB88) also hypothetically compensates for the missing Belgian data by its negative bias. The GB88 data had been collected on a different sampling rational: probability sampling in contrast to quota sampling in the Eurobarometer. This provides a unique opportunity to test the difference of these two sampling methods for the British case. We know from Miller's analysis (1990a) that the quota sample yields higher results than the probability sample, because of the self-selection bias of quota interviewing. We tested that on the 8-item PUS scale: the British score in the Eurobarometer was 4.58, in the GB88 survey it was 3.84. The difference of -.74 is significant (t=10.00; p=.0000; df=2971). We assume that this negative bias lowers the EC mean and thus compensates for the missing Belgian data. In summary the size of the GB88 data compensates for the missing Danish and British data. The negative bias in the GB88 data compensates for the Belgian data for aggregate calculations. Our present analysis is based on a 10 EC country sample including Northern Ireland, which with good reason represents the 12 country EC. This data is used to make comparison to the combined US data from 1988 and 1990.

Second, in order to calculate Jon Miller's literacy index (Miller, 1983) the re-integration of the new data into the old data is necessary, once the recoding is completed. However, it was not possible to match the case numbers of the verbatim responses with the case numbers in the old data file. The case numbers on the new data did not match the case numbers in the old file. In the course of the data cleaning Faits et Opinions has transformed the national level case numbers into a European level case number in a way that is impossible to reconstruct. Single cases have gone missing, and it is impossible to find out which number these cases had on the national level. Non identical cases would be matched, and the resulting new data file would be invalid. Hence, we confine our discussion to a general critique of the 'literacy' and 'attentive public' index. We will instead focus on the development of an alternative coding frame that allows us to map successfully the qualitative diversity of PUS around Europe.

4.2 The American approach: 'attentive public' and 'science literacy'

Millers' **attentive public** is defined by crosstabulating a) items which measure high **interest** in either science or technology, b) at least moderate **self-reported informedness** in either issues, and c) regular **newspaper readership**[9]. This measure shows a significant difference between the EC (19%) and US (21%). The error margin of the two samples is 1.91%. The Americans are on average more attentive to science and technology than the Europeans (Miller in Durant et.al, 1991). Methodological problems with Miller's definition of the attentive public, particularly in relation to time series data, have been pointed out by Beveridge and Rudell (1988): changing definitions and relaxation of requirements. In order to establish good time series data, a consistent definition of variables is a condition sine qua non. Because of the problems with the data matching we are unable to replicate Miller's '**scientific literacy**' index for Europe so that it can be compared accurately.

Miller's index is ambiguously defined as will be shown. It appears that Miller applies it in a flexible way to different data contexts. A procedure which has at the same time its advantages and disadvantages. The advantage is that the approach can be flexibly applied to various different data sets to open the debate on the issues, and give tentative cross-national comparisons. The disadvantage is that, strictly speaking, the comparisons are inaccurate. For accurate and precise comparisons the survey data need to be based on exactly the same items, the same data sampling method, and the same analytic procedures. Jon Miller's agenda was and still is to call attention to the issue of public understanding of science in a multi-national context, and to the crisis of science education in the USA. We will provide a technical critique of Miller's approach, while acknowledging his success in setting up a forum for the international comparison of data and setting the agenda as a consistent pioneer of the field[10].

According to Miller (1991) 'scientific literacy' comprises a) a basic vocabulary, b) an understanding of the process and the methods of science, and c) an understanding of the impact of science and technology on society. Miller defines 'literacy' as a

---

[9] In Durant et.al. (1991) Miller defines the attentive public as follows: 'If the respondent reports that he or she is very interested in and very well informed about either or both of these issues and indicates a regular pattern of newspaper or magazine readership, then the respondent is classified as attentive to science and technology. In addition, those respondents who report that they are very interested in both new scientific discoveries and new inventions and technologies, who classify themselves as moderately well informed on both issues, and who are regular readers of newspapers or magazines' (9)

[10] Jon Miller was one of the main initiators of the ICCSPUST (International Council for the Comparative Study of the Public Understanding of Science and Technology), an international network that was set up to meet annually and to coordinate survey efforts on science and technology for purposes of international comparison. Meetings have been held in London (May 1990), in Washington (Feb 1991), and in Tokyo (Oct 1992).

threshold measure; a person is literate or not. The measure is operationalized by tabulating surveys data across a number of variables. He does not provide, however, a consistent measure of literacy. He flexibly constructs the index in each survey on the basis of different items, notably within the above framework. This is clearly shown in the differences between the measures deployed in 1983 and those deployed in 1991, as the following example shows.

> In 1983 '**process**' is defined by the open question and two astrology items. The '**scientific vocabulary**' is based on radiation, GDP and DNS items. The understanding of the **impacts of science** is measured by a rating of the costs and benefits of controversial scientific programs like 'nuclear power, food additives, and the US space program'.
>
> In 1991 'scientific vocabulary' is defined by an arbitrary 66% correct answers on a **9 item scale of factual knowledge**. 'process' is defined by an 'acceptable answer' to the open question, without specifying the 'acceptable answer', and the astrology question. **Astrology is taken as a demarcation variable.** Whoever perceives some kind of scientific virtue in astrology is disqualified from 'true' scientific understanding.

It remains unclear how the 'literacy' index is calculated and should be calculated in the future. A transparent statistical procedure is substituted by the vague criteria such as 'a minimal acceptable score in all dimensions' (Miller 1991, 41).

The demarcation of science and non-science with the help of astrology is problematic. Miller uses a question on astrology 'Could you say that astrology is very scientific, sort of scientific, or not at all scientific?' as a measure of threshold of scientific literacy. People responding 'very.. or sort of scientific' are considered scientifically illiterate. Our analysis of European data shows that beliefs in astrology and scientific understanding tend to be in a inverted U-type, non-linear relationship (Durant and Bauer, 1992). People with a medium level of public understanding of science (PUS measure) most strongly uphold the scientific status of astrology. Acknowledging scientific status for astrology may well reflect an embryonic interest in and understanding of science, rather than its demise. The popularity of astrology among certain sections of the public rather reflects a crisis of religion and a consequence of social disintegration in modern liberal society, than a crisis of the scientific culture (Boy, 1992; Durant and Bauer, 1992). It seems inadequate to take the belief in astrology as a cut off point.

To summarise, the data shows that the Americans are on average more attentive to matters of science and technology. The concept of 'scientific literacy', however, is not consistently defined. We recommend extreme caution, particularly in relation to times series data, with any measure of the concept of 'scientific literacy'. Based on the European evidence we regard astrology as misplaced as a demarcation criteria and threshold measure for the definition of scientific literacy.

4.3 Problems with ranking the nations on PUS

The simplest approach to comparing different nations is to rank them on a number of scores. Our analysis uses the facts score and the process score as shown in table 3 (see appendix 7.6; table 8 and 9 for national scores). The process score is the coding frame used for the US responses to the open question. We have applied several different definitions of the facts score, 10 items[11], with 8 items and with 12 items, in order to show clearly the vicissitudes of the ranking of nations. From table 3 we can reach several conclusions: First, different criteria rank the nations differently. The most volatile are the British and the Germans; depending on the items included, they will score differently. Including the 'radioactivity' and the 'continental drift question' in facts(10) puts the Germans five ranks back, from the top to rank 6, and the British three ranks up from 5 to rank 2. Including all the questions puts the British at the top. Second, the ranking is unstable only for the top group of nations. The differences among the high level group are very small, and not significant (see appendix 7.7; table 7). It is statistically nonsense to rank scores according to the empirical means, because differences may be due to measurement error. The estimate of error of mean scores is .17 for the 8 item scale (t=1.96; s=1.9; n=1000) and .21 for the 10 item scale (t=1.96; s=2.35, n=1000). The minimal difference between two scores has to be larger than the margin of error to justify a ranking. Grouping the 12 EC nations into high, medium and low level of PUS, provides a more stable classification across several indicators.

[11] The facts score with 10 items is from the US Science and Engineering Indicators (NSB, 1991). Our own calculation result in slightly different scores. The origin of that diversion is unknown. The scores of Spain, Netherland, Great Britain, and Denmark are higher in the calculations of the US Science Indicator than in ours. The rank of Britain would fall from 2 to 4 in our calculations. The error affects the ranking, but does not affect the clustering of nations into three groups. We will use the US data, despite that doubt about the calculation, because is has been granted official publication in the USA, and will therefore be widely used as an indicator. The 8 item score is identical with the above score but takes out those items which have different wording in the American and in the European survey: the question on the continental drift and the question on man-made radioactivity. The 12 item score has been used in an earlier analysis of all the European items disregarding the comparability with the American survey (Bauer et.al., 1991).

Table 3: The ranking of nations in PUS

Nation	facts(10)	facts(8)	facts(12)	Process
High level				
Denmark	1	2	5	-
UK [12]	2	5	1	9
Luxembourg	3	3	3	8
France	4	4	2	4
Netherland	5	6	4	1
Germany	6	1	6	3
	US90	US90		
Medium level				
Italy	7	7	7	2
				US90
Belgium	9	9	9	-
Eire	10	10	8	11
Low level				
Spain	11	12	11	10
Greece	12	11	10	12
Portugal	13	13	12	6

'Process' is measured by the open question coded with the US coding frame

Third, the ranking of factual scores and of process scores in very different. The US type coding of the open question puts Britain and Luxembourg into the middle group. Italy and Portugal are put into the top group of understanding of scientific method. Southern European countries tend to have more missing values and a **bimodal distribution of scientific knowledge** (Bauer et al., 1991). The 'elite' in Mediterranean countries is responding to the open question in a way that scores high on the US coding frame, and many people do not respond at all. We excluded the missing values from the analysis. That procedure introduces a positive bias into the analysis, particularly in the case of Portugal (missing values=48% in contrast to Netherland missing=18%). Fourth, the USA is ranking consistently between the high level group and the medium level group. In the understanding of scientific method the US ranks in the middle group. The average scores for US and EC can be compared. The EC tends to be lower on average

[12] The data for the UK does not include the data for Northern Ireland. Eurobarometer collects data for Northern Ireland and Great Britain in different samples. Combining the two data sets for UK requires adequate weighting procedures. We will not do that here. For the present purposes not the precise value of the UK or its position in the ranking is important, but the fact that such a position depends on the kind of measure taken, even within the same survey.

in scientific understanding, although the significance of the difference in unclear[13].

In summary: ranking nations on PUS scores is **statistically inadequate** because a) the differences, particularly in the top group of PUS, are not significant, and b) the rankings depend essentially on the kind of items used. Comparing the US and the EC is possible. The EC tends to be less understanding in matters of science and technology than the USA on average. The statistical significance of this result is not clear. We recommend to group nations into three clusters, which yield robust results across different measures, rather than to rank nations.

[13] The difference reported in the US Science Indicators 1991 is not significant. The error margin for means in US and EC comparison on a 10 item scale is .07 (US pus=5.79; EC pus=5.75). However our own calculations reveal a significant difference (US pus = 5.79; EC pus=5.47). The origin in the differences in the calculation is unclear.

4.4 Towards a map of European Public Understanding of Science

4.4.1 The approach of the Science Museum, London

The approach taken by the group at the Science Museum London and its British collaborators (Durant et al., 1989; Evans et.al, 1989; Bauer et al., 1991) can be characterized by the following points:

a) We base our discussion of the Public Understanding of Science on scalar measures, which are tested for their internal consistency. This element of our approach has been taken over by the US National Science Board (1992)

b) Instead of looking only at average scores, we focus on the variance of PUS scores in a population and pay particular attention to unimodality or bimodality of that distribution. Bimodality may reflect an elite educational culture with regard to science and technology (Bauer et.al, 1991).

c) Rather than ranking national scores along insignificantly different average scores, that depend highly on the items included, we cluster national scores in three groups: high, middle and low.

d) We focus on the structural analysis of variables within nations to characterize different representation of science and technology among the public (Durant et.al, 1992). To compare such complex data structures across nations is our methodological aim.

e) We analyze the open question on scientific methods with the **Science Museum Coding Frame**, that subsumes the American intention, while it goes further at the same time. It allows us to measure the complexity of the answers, and eventually to map qualitatively different ways of understanding matters scientific among national elites.

4.4.2 The Science Museum Coding Frame

We have focused in our analysis of the Eurobarometer no 31 on the last point e). Our research team developed an alternative way of coding the open ended question. We take it that a good coding frame is as much theoretically informed as it is grounded in the data, we chose two theoretical ideas with our coding frame. They define its two dimensional structure:

The first dimension introduces a threefold distinctions that may be important in different 'understandings' of science and in mapping technology. First, an understanding that stresses scientific method (process); second, an understanding that views science as a social institution (institution); and third, an understanding that focuses on the effect of science and technology (effect). In addition, we code the kind of examples people give to illustrate their answer (example). The focus on method is the normative approach, which is often informed by **scientific teaching and textbooks** and philosophical discussions. It represents the self-conception of scientists. The focus on institutions is the **sociological approach**, which stresses the fact that science is an organisation among others, giving people a living and social position. The focus on effect is value oriented like a **utilitarian cost-benefit approach** or the modern approach of **impact analysis**. It basically tries to define science in terms of what are the costs and the benefits. We expect all three notions to appear in the public responses.

Second, sociological theory distinguishes at least three levels of systemic analysis: the interactional, the institutional, and the societal level (e.g. Luhmann, 1984). These levels are integrated and mutually constraining in their activity: what is possible on a higher level depends on the lower levels an vice versa. The interactional level describes the roles and expectations people take in their lives. The organisational level describes collective bodies and the kinds of technical expertise they create. The societal level describes generalized norms and culture. Culture may or may not be identical with a political nation.

Combining these two dimensions for process, institution and effect yields a simple grid of nine categories as described in our code book (see appendix). The first dimension defines the variable: process, institution, effect. The second dimension defines the variable value: interactional and role related, institutional, general norm of society. Six classes of examples are differentiated. The effects are further separated into positive and negative effects on each level. The total frame combines 21 categories - three for 'process', three for 'institution', six for 'effects', six for examples, 'complicatedness' as a combination of the four dimensions, do not know and no response - to code the content of the responses to the open question. The American frame was restricted to only five categories.

Table 4: The Science Museum Variables

dimensions	number of subcategories	level of measurement
Process	3	ordinal
Institution	3	nominal
Effect	6	nominal

Example	6	nominal
Complicatedness	1	ordinal
do not know 1		
no response	1	
Total	21	

The additive combination of all categories gives a measure of the **complicatedness** of the answers[14]. Table 4 summarizes the dimensions, the number of subcategories for each dimension and the level of measurement, which are defined by the Science Museum Coding Frame, and gives the level of measurement. The Science Museum uses multi-dimensional codings to analyze the open question. Each response is individually coded on at least one of the variables. Several codings are possible. The Science Museum coding builds on the US coding frame, but takes the rest category, up to 35% of responses, of the US coding frame as an important source of further information about people's notions of science.

4.4.3 The Reliability of Science Museum Coding

The reliability of the Science Museum coding frame is better than the American version as shown in table 5.

Table 5: The reliability of the SM coding frame

variable	kappa	
US coding	.44	American coding frame

[14] Complicatedness is defined by the sum of the dichotomized variables. Before the variables are added, each variable is recoded into 0 or 1: 1 = the category is coded; 0 = the category is not coded. Missing variables are included as 0.

We use the word 'complicatedness' instead of 'complexity' to make an important distinction: sets can be ranked by the number of elements (complicatedness) and the number of relations among the elements (complexity). Sets of equal complicatedness can differ in complexity depending on which relations among the elements are actualized. Complicatedness refers to the potentiality, while complexity describes the actuality of relations. Our measure is a sum of elements and can therefore not a proper measure of complexity. To measure the complexity of the answers we would have to further analyze how the different elements of process, institution, effect and examples actually are combined. We owe the attention to this problem to a discussion with Angela Stathopoulou.

Process	.62	Science Museum Coding Frame
Institution	.55	"
Effect	.40	"
Example	.68	"
Complicatedness	.42	" "

The variables Process, Institution and Example are more reliable than the simple US coding. The variables Effect and Complicatedness are as reliable as the US coding frame. All codings are well beyond random judgements. The differences in reliability have to be taken into account for the interpretation of the results. Results of the variable process, institution and example are more important than results from effect or complexity. The Science Museum Coding Frame is more reliable than the US coding frame, and more adequate for the data, although further improvements are necessary. Because the frame is more complex than the American one, coder training will be necessary to ensure reliability.

4.4.4 The Results of Science Museum Coding

The open ended question yields data only from a fraction of the population. Many respondents give very short answers, and many do not give any answer at all. Between 52% (Portugal) and 82% (Netherland) have responded to the open question, so that a coding is possible. On average about 75% of respondents give an answer. 65.4% of these responses have been coded for 'process' (Science Museum coding frame), 14.8% for 'institution', 12.1% for effect, and 7.7% for example (see appendix 7.7).

35% of the answers allow a coding that goes beyond a one-dimensional Popperian (theory, hypothesis testing, experiments, measurement) notion of science (definition: rest category on US coding frame). 20.2% of all codings are multiple codings and reflect more complicated answers to the open question (definition: the variable 'complicatedness' >= 2; see appendix 7.7). These responses are the answers of a 'minority' of the population[15]. We regard these answers as an important source of information, that is lost when we use a one-dimensional frame like the US coding frame. It will be essential to investigate the socio-demographic position of the those people that give answers that go beyond the mere methodological dimension of science. We would expect that giving such answers does not necessary correlate linearly with other measures of PUS and the level of education. Our failure to integrate the old and the new data prohibited the investigation of that question. This is clearly a task for a future survey.

In addition the Science Museum coding frame allows us to measure the **complicatedness** of the given answers. Complicatedness is a five point scale with a

---

[15] We cannot analyze the socio-demographics of the institutional, effect and example answers, because the data integration was not possible. This is indeed a very important research question for further surveys using the open question together with the Science Museum coding frame.

range from 0-4, 0 meaning no answer, 4 meaning a answer that combines all four elements of science: process, institution, effect and example. 21% or respondents combine two or more elements, 0.4% combine all four elements. The analysis of the complexity of the answers to the open question needs to be done in relation to socio-demographic and other variables. This is clearly a task for future research.

Rather than ranking the nations we follow a clustering approach. The appropriate method to integrate categorical data is a statistical method called 'correspondence analysis'[16] (Weller and Romney, 1990). Correspondence analysis is an analogue of multi-variate analysis methods for categorical data. It allows to approximate the structure of crosstabulated data into a geometrical representation, so that categories that covary are near to each other in a spatial representation. The results of our analysis are shown in figures 1 to 4. These figures map nations and coding categories into a hypothetical space of two dimensions, and show, by grouping the nations, which nations are similar and with respect to which category they are similar. It is a graphical way of interpreting crosstabulations.

In order to simplify the argument, we will not concern ourselves with the interpretation of the scales and the scalar values that represent each point on each figure. For the interpretation of the figure we look how the nations and the coding variables are grouped, and take that similarity between nations as our main results. Similarity is represented by the distance of two points. We can identify

a) how the nations are grouped, and

b) which variable characterizes each grouping best.

Because the figures are small some points cannot be properly labelled. Numbers mark overlapping points that can be identified in the legend that is given below the figures. The two dimensions define the cartesian coordinates for each point. The interpretation given here is tentative and based on the graphical representation and has the purpose to demonstrate the potential of the method of analysis rather than giving final results.

[16] We used the procedure ANACOR from the SPSS module 'Categories' (1990) for our analysis.

Figure 1: Variable 'process' by nation

The variables process has three values: theore = theory and hypothesis testing; exp = experimentation; depth = in depth investigation.

```
      ++-----+-----+-----+-----+-----+
   1.24+          theor               +
       |                              |
       |                              |
D      |                              |
i      |                              |
m    .68+                              +
e      |                              |
n      |                              |
s      |                F   NL        |
i      |    US88                      |
o    .11+      G                Lux    +
n      |       (1)                    |
       |  exp      It  (2)            |
2      |        NIr                   |
       |                       Gre    |
   -.45+                              +
       |                              |
       |        Eire                  |
       ++-----+-----+-----+-----+-----+
       -.86  -.42   .03   .47   .91
                 Dimension 1
```

SUMMARY OF MULTIPLE POINTS IN THE PLOT

POINT	DIM1	DIM2	ACTUAL LABEL OR NAME
(1)	-.14	.03	GB88
(1)	-.26	-.02	E
(2)	.34	-.06	depth
(2)	.41	-.15	P

Figure 1 groups the nations according to people's notion of scientific method in the narrow, popperian sense, the variable 'process'. The **US, Britain, Germany, Northern Ireland, Spain and Italy** are similar and group around the idea of experimentation. No nation seems to associate science very closely with theory. The closest to theory are **France, Netherlands, the US, Germany and Luxembourg**. The notion of 'in-depth' investigation is characteristic for the **Greeks and the Portugese. The French, the Dutch and the Luxembourg** are tend rather to a notion of 'in-depth' investigation rather than experimentation, when asked about a scientific study. The Irish have a notion of scientific method that apart from the other nations. It is surprising that the pattern does not provide clear north/south or Anglo-Saxon/Latin division. Things seem to be more complicated, indeed.

Figure 2: Variable 'Institution' by nation

The variable 'institution' has three values: norm = a common, societal and normed enterprise; name = naming an organisation or an individual; gen = the general role, image of the scientist.

```
       ++-----+-----+-----+-----+-----+
    .69+                        E    +
       |                        P    |
       |Eire                         |
D      | GB88          G    gen      |
i      |   name                      |
m   .04+             It              +
e      | (1)                Gre      |
n      |               NL            |
s      |                             |
i      |                             |
o  -.61+            Lux              +
n      |                             |
       |                             |
2      |           F                 |
       |            norm             |
  -1.25+                             +
       |                             |
       ++-----+-----+-----+-----+-----+
     -1.04  -.53  -.02   .49  1.00

                 Dimension 1
```

SUMMARY OF MULTIPLE POINTS IN THE PLOT

POINT	DIM1	DIM2	ACTUAL LABEL OR NAME
(1)	-.77	-.03	US88
(1)	-.79	-.08	NIr

Figure 2 groups the nations together according to their notion of science as a social institution. A clear structure emerges, that is different from the first one. **Ireland, Britain, the US and Northern Ireland** focus on science as a particular organisation, like laboratories, universities or foundations. The **French and to some extent the Luxembourg** see it as a common societal enterprise. The **Spanish, the Portugese and the Greek**s focus on the individual role of 'the scientist' and his or her characteristics. The Germans, the Italians and the Dutch are similar in being not distinct on the institutional category.

Figure 3: Variable 'Effect' by nation

The variable 'effect' has six values: +pers = positive personal outcomes; -pers = negative personal outcomes; +loc = localized positive outcomes; -loc = localized negative outcomes; +hum = progress of human kind; -hum = damage or threat to human kind.

```
       ++-----+-----+-----+-----+-----+
   1.08+           P                  +
       |         F   +hum E           |
       |                              |
D      |                              |
i      |                              |
m   .33+   NIr(1)                     +
e      |                              |
n      |  GB88                        |
s      |      +loc     NL        Gre  |
i      |                              |
o  -.42+    -hum           (2)        +
n      |                 -pers        |
       |                              |
2      |           G                  |
       |                              |
  -1.18+Eire                          +
       |-loc                          |
       ++-----+-----+-----+-----+-----+
      -1.26  -.67  -.08   .51  1.11
```

SUMMARY OF MULTIPLE POINTS IN THE PLOT

POINT	DIM1	DIM2	ACTUAL LABEL OR NAME
(1)	-.49	.38	US88
(1)	-.26	.37	It
(2)	1.03	-.37	+pers
(2)	.78	-.49	Lux

Figure 3 groups the nations according to their focus on effects. Negative effects are important for the **Irish and the Germans**. The Irish see negative effects of particular areas of science. The Germans tend more towards negative outcomes for the person and for humanity in general. The **French, the Portugese and the Spanish** see science as a blessing for humanity. For the **US, The Northern Irish and the British** positive outcomes of specific areas prevail, albeit the British are inclined to associate science with threats to humanity as a whole. Luxembourg and Greece are concerned with the personal outcomes of science, positive and negative. The Dutch are not distinct with regards to the effects of science.

Figure 4: Variable 'Example' by nation

The variable 'example' has six values: med = medical and pharmaceutical examples; life = other life sciences; phys = physical sciences; tech = technical sciences and technical objects; soc = social sciences; oth = other examples.

```
       ++-----+-----+-----+-----+-----+
    .51+  P    E  Eire              +
       |     (1)     F       life    |
       |    (2)                NL    |
D      |      phys (3)               |
i      |                             |
m  -.36+                             +
e      |        G                    |
n      |                             |
s      |                             |
i      |          US88               |
o -1.24+                             +
n      |                             |
       |                             |
2      |                             |
       |                             |
  -2.12+                             +
       |                             |
       |           soc               |
       ++-----+-----+-----+-----+-----+
     -1.24  -.55   .14   .83  1.52
                 Dimension 1
```

SUMMARY OF MULTIPLE POINTS IN THE PLOT

POINT	DIM1	DIM2	ACTUAL LABEL OR NAME
(1)	-.54	.37	oth
(1)	-.44	.30	Gre
(1)	-.46	.27	tech
(2)	-.64	.17	NIr
(2)	-.30	.12	It
(2)	-.34	.09	GB88
(3)	.09	-.06	med
(3)	.45	-.09	Lux

Figure 4 groups the nations according to the examples that were given by the respondents. Technical, medical and examples from the physical sciences are characteristic for most countries. Clearly distinct are the **Netherlands**, which gives examples from the life sciences other than medicine, and **Germany and the US**, which are inclined to associate examples from the social sciences.

4.5 Comparing US and Europe

In table 6 we summarize the results for the US and Europe on five variables: attentive public in the Miller definition, process measured in both versions, with the US coding frame and with the Science Museum coding frame, and the variables institution, effect, examples and complicatedness.
Error margins are given for the comparison of the US and the European average scores. These margins are based on unequal sample sizes as shown in appendix 7.6 (table 7).

Table 6: Comparing US and Europe

Variable	US88	US90	Europe	margin 95%[17]
Attentive Public	--	21%	19%*	1.91%
Process (US code)*	3.46	3.52	3.58*	.067
Process (SM)*	2.43	2.40	2.61*	.051
Institution	10.8%	8.7%	7.3%	4.5%
Effect	5.5%	4.5%	8.2%	4.6%
Examples	12.3%	11.5%	6.9%	5.9%
complicatedness*	.85	.81	1.08	.018

* = significant at 5 % alpha error

'Process' is measured in a way that the lower value expresses the prevalance of a Popperian notion of science: the lower, the more Popperian.

Four conclusions on the comparison of Europe and the US can be drawn: First, the **US is more attentive** to issues of science and technology than Europe on average. Second, the **Americans have a more popperian understanding** of science than the Europeans (process EC=3.58; US=3.46 or 3.52). Third, Americans and Europeans give on average **similar answers to the open question**: institution, effect, examples. The differences along these variables are not significant. Fourth **Europeans give more complicated answers** than Americans. European do more often combine process, institution, effect and examples in their answers to the question 'what does it mean to study something scientifically'.

On the whole we conclude that comparisons across continents is not very fruitful. To look at the structure of these variables across nations is more rewarding as has been

[17] Error margins are calculated for a 5% alpha error: t = 1.96.

shown in the previous chapter. Depending on the variable the US will group with different European countries. With respect to process understanding, the US is similar to Germany and Britain. With respect to science as a social institution, the US is similar to Britain and Ireland. With respect to the perception of effects, the US is similar to Britain and Northern Ireland.

With respect to scientific examples given, the US is most similar to Germany. The old Anglo-Saxon connection is clearly emerging in the similarity of responses between the US and Britain.

5. Recommendations

The measurement of Public Understanding of Science

1. to add a **third dimension** of public understanding of science in future surveys: 'knowledge about the social organisation of science' (science policy, science enterprises, laboratories, peer review system). Adequate items are to be developed.

2. We recommend to group nations into three clusters: high, medium and low, which yield robust results across different measure, rather than to rank nations along insignificant differences.

The measurement of scientific literacy

3. The concept of 'scientific literacy', however, is ambiguously defined. We recommend great caution, particularly in relation to time series data, with any measure of the concept. Based on the European evidence we regard the issue of astrology as misplaced in the definition of scientific literacy.

The handling of the open question

4. With a different form of analysis the open question proves an important source of information for mapping the cultural and other differences among nation. The open question must be included in further surveys. However more care has to be given to the translation of the question, the transcription and coding of verbatim responses.

5. The translation of the open question has been incorrect in four languages (France, Italy, Greece, Netherland). This has introduced a significant translation effect, that positively biased the results of these countries in the open question. It is necessary to give more care to the correct translation of the question.

6. The translation of the open question needs to be standardized in the future. We recommend the personalized version, because it seems easier to handle from the respondents' point of view. The two versions may lend themselves to a split half design.

7. It may be desirable to have a standard format for the presentation of the verbatim responses to the open question.

8. DG 12 should take clear responsibility for archiving the verbatim responses to the open question in the future. For the open question a separate report is necessary, which contains the verbatim responses of each respondent, as it was provided by the Spanish, Dutch and Portugese agency in 1989.

The coding of the open question

9. We recommend the introduction of **transparency and standardization** into the transcription and coding procedures of the open question in the following order of priority:

- A typed record of the verbatim responses is needed
- The coding frame as used needs to be published
- Interviewer and coder have to be different persons
- The interviewer must not be aware of the coding frame. If he or she is aware of it, that needs to be said.
- The coding procedure needs documentation
- Reliability indices must be calculated for the coding

10. We recommend to clarify the definitions of the American categories as well as pursuing alternative methods of analysis, particularly in the European multilingual context.

11. We offer a revision of the coding frame for the European context mainly for two reasons: a) to accommodate the different 'natural' meanings of 'scientific method' in different languages, and b) to match the sophistication of responses.

12. It is necessary to relate the socio-demographic position of the those people who give answers to the open question that go beyond the methodological dimension of science. We would expect that giving such answers does not necessarily correlate linearly with other measures of PUS or with the level of education. Our failure to integrate the old and the new data prohibited the investigation of that question. This is an important task for a future survey.

6. References

Bauer, M. and J. Durant (1991) Surveys in Public Understanding of Science: a statement of principles and an agenda for further research, an unpublished statement to the National Science Foundation, Washington, London, July 1991.

Bauer, M., J. Durant, and G. Evans (1991) European public perception of science: an international comparative study. Paper presented at the National Conference of the AAAS, Washington, February.

Beveridge, A.A. and F. Rudell (1988) An Evaluation of 'public attitudes toward science and technology' in Science Indicators: the 1985 report, Public Opinions Quarterly, 52, 374-385.

Boy, D. and G. Michelat (1992) The French and the pseudosciences, Paper presented to the Annual Meeting of the AAAS, Chicago, February, 1992

Commission of the European Communities (1990) Eurobarometer: what it is and how to use it, Brussels, November, 1990.

Durant, J., G.A. Evans and G. Thomas (1989) The public understanding of science, Nature, 340, 6 July, 11-12

Durant, J.R.; J.D. Miller; J.F. Tschernia and W.vanDeelen (1991) Europeans, Science and Technology, Paper presented at the National Conference of the AAAS, Washington, February.

Durant, J.; G. Evans, and G. Thomas (1992) Public Understanding of Science in Britain: The role of medicine in the popular representation of Science, Public Understanding of Science, 1, 2, 1992.

Durant, J. and M. Bauer (1992) British Public perceptions of astrology. An approach from the Sociology of Knowledge. Paper presented to the Annual Meeting of the AAAS, Chicago, February, 1992

Eurobarometer no 31 (1989) Pubic Opinions in the European Community, Brussels.

Evans, G. and J. Durant (1989) Understanding of science in Britain and the USA, in: British Social Attitudes - special international report, London, Gower.

Harre, R. (1972) Philosophies of Science, an introductory survey, Penguin, Harmonthworth, 1989.

Krippendorff, K. (1980) Content analysis, Beverly Hills, Sage Publications.

Luhmann, N. (1984) Soziale Systeme: Grundriss einer allgemeinen Theorie, Frankfurt, Suhrkamp.

Miller, J.D. (1983) Scientific literacy: a conceptual and empirical review, Daedalus, Spring, 29-48.

Miller, J.D. (1990) Note on the effect of quota and probability sampling on selected scores, comparing EC data and GB88 data, unpublished manuscript, DeKalb, US

Miller, J.D. (1991) The Public Understanding of Science and Technology in the US, 1990, a report to the National Science Foundation, Public Opinions Laboratory, Northern Illinois University, DeKalb.

National Science Board (1992) Science and Engineering Indicators, Washington

Popper, K.R. (1963) Conjectures and refutations. The growth of scientific knowledge, London, Routledge and Kegan Paul.

Suchner, R.W. (1985) The attentive public for organized science. A summary of findings, presented to the annual meetings of the Midwest Sociological Society, St Louis, Missouri, April 12, 1985.

Wellner, S.C. and A.K. Romney (1990) Metric scaling. Correspondence Analysis, Quantitative Applications in the Social Sciences no 75, Newbury Park, Sage Publication.

Withey, S.B. (1959) Public Opinions about Science and the Scientist, Public Opinions Quarterly, 23, 382-388.

7. Appendix

7.1 The coding frame as reported by 'Faits et Opinions' 1989

7.2 The coding frame reported by Gallup GB in 1989

7.3 The coding frame used for the US and GB data in 1988

7.4 The Science Museum Coding Frame

7.5 The Items of the Knowledge Scales

7.6 Selected data tables

7.7 The distribution of the SM coding variables

7.1 The coding frame as reported by 'Faits et Opinions' 1989

These are the variable labels from the raw data set:

1. theories and hypothesis

2. experiments, tests

3. exact examination

4. measure, classify

5. other

6. Does not know

These are the variable labels, as they are reported in the computerized data set. It remains unclear how the different coding frames, like the one used by Gallup GB, is integrated into the above coding frame.

7.2 The coding frame reported by Gallup GB in 1989

1. Look into/examine/study all facts/data/aspects. Analyze, take all factors into account

2. In depth/detail. Delve into. Detail analysis.

3. Without prejudice/unbiased/objectively etc.

4. Examine/look into by experts/scientists. Examine scientifically using scientific knowledge/evidence

5. Monitor/study/measure/record results through each stage of development

6. Scientific tests/experiments. Scientific theory worked out in practice/testing of theory/controlled test

7. Read/study literature about it

8. Examine under microscope/laboratory conditions. Laboratory work/using technology/modern technology

9. Research

0. None/nothing/not a lot really/nothing much/no real impression
X. Other

V. Does not know/no idea

7.3 The coding frame used for the US and GB data in 1988

Priority code in order listed below (code 1 has highest priority).

1. Theory construction and testing

    Response states that studying something scientifically means that is is studies in the context of a theory about the problem/phenomenon being examined, and/or that it is an attempt to disprove a hypothesis about the nature of the phenomenon/problem being studied.

    The words 'theory' and/or 'hypothesis' would almost certainly need to appear in the response to justify inclusion in this code.

2. to undertake experiments/tests

    Responses not falling into code 1 above which refer to the process of the study being to carry out experiments or tests in a strictly controlled way (this may be implied rather than explicitly stated). Words used, in addition to experiment or test, could be 'using strict controls', 'control groups'.

3. Open-minded, rational in-depth exploration of phenomena/problem to be examined

    Responses to not fall into codes 1 or 2 but which talk about evaluating the problem in an unbiased/open minded way, taking into account all possible information, studying it on a rigorous (logical basis).

4. to measure or classify but no mention of any rigour in the process

    Codes 1-3 do not apply to response. Response may describe study in terms of concrete actions used by scientists (e.g. use a microscope or telescope) or it may talk about measuring or classifying but without mentioning the need to use an unbiased rational approach.

5. other answers

    (except those falling into codes 8 or 9 below)

8. Does not know/guessed etc.

9. not answered

7.4 The Science Museum Coding Frame

The coding frame uses multiple coding on four dimensions

process (first digit) process

1 - in depth analysis, discovery, attitudinal
2 - measurement, mathematics, experiment
3 - hypothesis and theory testing

7 - not mentioned
8 - do not know
9 - missing

institution (second digit) instit

1 - general role and image
2 - naming and organisation or an individual
3 - common societal and normed enterprise

7 - not mentioned
8 - do not know
9 - missing

effect, outcome (third digit) effect

1 - positive personal growth
2 - negative personal outcomes

3 - localized positive outcomes
4 - localized negative outcomes

5 - Progress of human kind
6 - threat or damage to human kind

7 - not mentioned
8 - do not know
9 - missing

example (forth digit) exam

1 - medical
2 - other life sciences
3 - physical sciences
4 - technology
5 - social sciences
6 - others

7 - not mentioned
8 - do not know
9 - missing

complicatedness

a combination of the four previous dimensions.

Complicatedness is defined by the sum of the dichotomized variables. Before the variables are added, each variable is recoded into 0 or 1: 1 = the category is coded; 0 = the category is not coded. Missing variables are included as 0.

coding conventions:

7777 other category

any answer which cannot be fitted into the category system

8888 'do not know'

multiple coding: a code has 4 digits, each digit corresponding to one of the four categories.

9999 'missing value'

no answer given to that question

1737

1: in-depth analysis
9: institutional dimension does not apply
3: a specific positive outcome is mentioned
9: no example is given

- in case several categories apply from one code three rules are used:

a) the most elaborated point is coded
b) the first mentioned is coded
c) the highest code possible in case of 'process'

Comment: A detailed Coder Handbook, that defines the categories and gives examples for each coding, is in preparation (Martin Bauer, Research Fellow, The Science Museum, London)

7.5 The Items of the Knowledge Scales[18]

- The 8 item scale

1. The centre of the earth is very hot (83.1%)
   is true is falsedon't know

2. The oxygen we breath comes from the plants (77.3)

3. Electrons are smaller than atoms (39.5%)

4. It is the father's gene which decides whether the baby is a boy or a girl (47.4%)

5. The earliest humans lived at the same time as the dinosaurs (43.5%)

6. Antibiotics kill viruses as well as bacteria (23.4%)

7. Lasers work by focusing sound waves (33.4%)

8. Radioactive milk can be made safe by boiling it (60.9%)

- Added for the 10 item scale

9. The continents are moving slowly about on the surface of the earth (64.2%)

10. All radioactivity is man-made (53.1%)

- Added for the 12 item scale

11. Does the earth go around the sun or does the sun go around the earth ?
    - the earth goes around the sun (80.0%)
    - the sun goes around the earth
    - don't know
    if correct then go to 12:

12: How long does it take for the earth to go around the sun ?
    - one day
    - one month
    - one year (50%)
    - don't know

---

[18] **All questions have been use in EC barometer no 31, 1989; For the analysis dk answers are coded as missing values; the percentage of correct answers give an index of the difficulty of the item, a good scale has a balanced mix of easy and difficult questions.**

7.6 Selected data tables

Table 7: number of observations per variable

variable	US88		EC89	
	sd	n	sd	n
UScode	1.35	1623	1.27	7182
Process	.63	1339	.61	6062
Complicated	.61	2074	.75	9184
Institution	---	200	---	1580
Effect	---	88	---	1450
Example	---	125	---	778

The differences in the number of observation are due to variations in missing values

Table 8: Mean value on various Knowledge scales

nation	8 items	10 items*	12 items
France	4.59	6.11	7.29
Germany	4.89	5.97	7.17
Italy	4.40	5.66	6.80
Spain	3.87	5.03	5.66
Portugal	3.45	4.09	4.84
Netherl	4.56	6.05	7.26
Luxembour	4.73	6.17	7.29
GB	4.58	6.17	7.35
N Ireland[19]	4.16	5.27	6.43
Eire	4.08	5.19	6.36
Greece	3.89	4.71	5.68
Belgium	4.08	5.24	6.28
Denmark	4.82	6.23	7.24
US90	----	5.79	----
EC	4.30	5.75	6.56
Alpha	.64	.71	.73
sd error	.17	.21	

* the 10 item scale is taken from NSB (1992, 469). Our own calculations have shown slight deviations from these figures: Spain = 4.85; Netherland = 5.98; GB = 6.07; EC = 5.47; for the items see appendix 7.5

[19] For survey technical reasons Northern Ireland (n=300) is covered with a seperate sample than GB (n=1000). The weighting information was not available at the time of completion of this report. We therefore show the results seperatetly.

Table 9: Mean values for the codings of the open question

nation	UScode	SM process	Complicatedness
France	3.36	2.62	.97
Germany	3.36	2.48	1.15
Italy	3.24	2.58	.81
Spain	3.77	2.53	.70
Portugal	3.46	2.69	.64
Netherl	3.18	2.68	1.40
Luxemb	3.51	2.76	.93
Eire	3.86	2.60	1.00
GB88	3.72	2.55	1.28
N Ireland	3.50	2.55	.81
Greece	4.11	2.81	.92
US88	3.46	2.43	.85
US90	3.52	2.40	.81
EC89	3.58	2.61	1.08
kappa	.44	.62	.42

The correlation between UScode and process is r = .39; the correlation between UScode and complicatedness is r = .07 and between Process and complicatedness r = -.03.

The wording of the question is: Please tell me in your own words, what does it mean to 'study something scientifically' ?

Process measures are coded in a way that a low value represents a 'Popperian' view of science, a high value a non-Popperian view: the lower, the more 'Popperian'

Table 10: Percentages of responses on three SM variables

nation	institution	effect	example
France	13.0	7.1	10.1
Germany	14.5	9.5	7.6
Italy	3.0	9.7	1.2
Spain	4.2	3.0	4.4
Portugal	6.2	4.2	2.4
Netherl	9.6	23.6	9.0
Luxembour	1.6	3.7	2.1
Eire	10.5	4.7	11.4
GB88	9.0	11.6	23.9
N Ireland	1.2	0.8	2.0
Greece	7.7	12.0	2.3
US88	10.8	5.5	12.3
US90	8.7	4.5	11.5
EC89	7.3	8.2	6.9
kappa	.54	.40	.67
n	1963	1611	1020
N	13291	13291	13291

kappa is the reliability index for each category; n is the number of coded answers = 100%; N is the total sample size.

When calculating the percentage, missing values are excluded; the 100% basis is the sum of all codings of the open question in EC89, GB88, US88 and US90. The percentages depend on the sample size.

7.7 The distribution of the SM coding variables

Process: US coding frame

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	515	3.9	5.0	5.0
	2	1954	14.7	18.8	23.8
	3	2861	21.5	27.6	51.4
	4	1392	10.5	13.4	64.8
	5	3648	27.4	35.2	100.0
	.	2921	22.0	Missing	
		-------	-------	-------	
	Total	13291	100.0	100.0	

Process: Science Museum Coding frame

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	531	4.0	6.1	6.1
	2	2817	21.2	32.4	38.5
	3	5343	40.2	61.5	100.0
	.	4600	34.6	Missing	
		-------	-------	-------	
	Total	13291	100.0	100.0	

Institution

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	728	5.5	37.1	37.1
	2	898	6.8	45.7	82.8
	3	337	2.5	17.2	100.0
	.	11328	85.2	Missing	
		-------	-------	-------	
	Total	13291	100.0	100.0	

Effects

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	429	3.2	26.6	26.6
	2	30	.2	1.9	28.5

3	689	5.2	42.8	71.3
4	74	.6	4.6	75.9
5	367	2.8	22.8	98.6
6	22	.2	1.4	100.0
.	11680	87.9	Missing	
	-------	-------	-------	
Total	13291	100.0	100.0	

Examples

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	1	394	3.0	38.6	38.6
	2	138	1.0	13.5	52.2
	3	153	1.2	15.0	67.2
	4	177	1.3	17.4	84.5
	5	42	.3	4.1	88.6
	6	116	.9	11.4	100.0
	.	12271	92.3	Missing	
		-------	-------	-------	
	Total	13291	100.0	100.0	

Complicatedness

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	.00	3190	24.0	24.0	24.0
	1.00	7410	55.8	55.8	79.8
	2.00	2244	16.9	16.9	96.6
	3.00	389	2.9	2.9	99.6
	4.00	58	.4	.4	100.0
		-------	-------	-------	
	Total	13291	100.0	100.0	

```
COUNT   VALUE

3190     .00  |_________________________:_
7410    1.00  |________________________________________________________:_______
2244    2.00  |____________________    .
 389    3.00  |_:
  58    4.00  |
              I.........I.........I.........I.........I.........I
              0      1600      3200      4800      6400      8000
                     Histogram frequency
```

Mean	1.000	Median	1.000	Std dev	.752
Skewness	.659	Minimum	.000	Maximum	4.000