NUSAP - The Management of Uncertainty and Quality in Quantitative Information

Dr. Jerome R. Ravetz, The Research Methods Consultancy, 111 Victoria Road, Oxford OX2 7QG, jerome-ravetz@tiscali.co.uk www.jerryravetz.co.uk
Silvio O. Funtowicz, E.C. Joint Research Centre, Ispra, Italy

I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.

(Lord Kelvin)

First, we must insist on risk calculation being expressed as distributions of estimates, and not as magic numbers that can be manipulaed without regard to what they really mean. We must try to display more realistic estimates of risk to show a range of probabilities. To help do this we need tools for quantifying and ordering sources of uncertainty and for putting them in perspective.

(W.D. Ruckelshaus)

Introduction

In this age of powerful information technology, decision-makers in all fields have instant access to masses of quantitative data to help them formulate their policies. These will usually be tested for their quality, by their success (or otherwise) in the achievement of their stated goals. But how much testing is there of the quality of the information on which the decisions are based? All over the industrial and administrative worlds, quality has become a keyword, and significant resources are devoted to quality-control of processes and procedures. Yet the quantitative information that is crucial to all such exercises has almost always escaped the same critical scrutiny.

Issues of uncertainty, and closely related, those of quality of information, are involved whenever research related to policy is utilized in the policy process. As these issues are new, we do not yet possess practical skills for dealing with them effectively. There are a variety of possible responses among users, contradictory among themselves, but all of them to be seen in ordinary practice. The simplest, and still most common, is when decision-makers and the public demand at least the appearance of certainty. The scientific advisers are under severe pressure to provide a single number, regardless of their private reservations. The decision-makers� attitude may be typified by the statement of a US Food and Drug Administration Commissioner, "I�m looking for a clean bill of health, not a wishy-washy, iffy answer on cyclamattes" (Whittemore, 1983). Of course, when an issue is already polarized, such simplicity does not achieve its desired ends. With such contested policy numbers in play, debates may combine the hyper-sophistication of scholastic disputations with the ferocity of sectarian politics. The scientific inputs then have the paradoxical property of promising objectivity and certainty by their form, but producing only greater contention by their substance.

It is easy to imagine, in a general way, what we mean by "quality" of information, on the analogy of the quality in any other input. Will it perform reliably at a reasonable cost? and, equally important, do we have good and proper evidence that it will do so? The absence of good quality is betrayed by high uncertainty. If we simply do not know, or we are not sure about the testing and certification of the input, then its quality is low, whether it be a plastic moulding or a set of numbers. Up to now, tests for the quality of quantitative information have been very undeveloped, in comparison with those in all other spheres. There are the standard statistical tests on sets of numbers in relation to an hypothesis; and there are highly elaborated formal theories of decision-making in which "uncertainty" is manipulated as one of the variables. But none of these approaches help with the simple question that a decision-maker wants to ask, when confronted with a screenful of information: is this reliable, can I use it safely?

There are two related reasons for the lack of methods in this important field. One is in our common assumptions about the way we know about the world. Science is based on numbers, therefore numbers are necessary for the effective study of the world; and we then assume that numbers, any numbers, are sufficient as well. In such an environment, it is not easy for anyone to imagine analyzing the uncertainties in numbers, and still less easy for him or her to get an audience. Although there is a rich literature of sceptical jokes about statistics, we still use them, usually quite uncritically, because there is nothing better to hand. One can even hear an argument for pseudo-precision in numerical expression (say, using three or four digits on very rough data), on the grounds that one has no idea of the uncertainties, therefore any "error bar" would be only a guess,

and so therefore let all the digits stand! If this policy were adopted in manufacturing, the consequences would be severe; are we sure that they are not equally so in finance and administration?

We have devised a notational system for quantitative information by which these difficulties can, to some extent at least, be overcome. It is based in large part on the experience of research work in the matured natural sciences. Contrary to the impression conveyed by popularisers and philosophers, the success of these sciences is not so much due to their being "exact" or certain, but rather to their effective control of the inherent uncertainties (of which inexactness is one sort) in their data and theories. Academics who teach and research in the sciences involved in decision-making, including economics and systems analysis, have found it difficult to resist "physics-envy", and are then hampered by an anachronistic and fantasized vision of the methods and results of physical science.

Of course, the myth of unlimited accuracy in quantitative sciences has many uses, principally for maintaining the plausibility of projects that have lost most or all contact with reality. For example, only with the debate over "Star Wars" did the public learn that computer programming is a highly uncertain and inexact art. With the proliferation of "models" of all sorts in natural and social science, we must recognize a new sort of pseudo-science, which we may call GIGO (Garbage In, Garbage Out). This can be defined as one where the uncertainties in the inputs must be suppressed, lest the outputs become indeterminate. Anyone whose work has suffered through reliance on low-quality quantitative information will have some idea of how much harm it can cause, and some inkling of how widespread is this particular form of incompetence in our society.

Remedying this situation is a big job; it will involve practitioners and teachers, and perhaps philosophers as well. In some ways the problem is built into our system of numbers, which was designed for counting and for simple calculation, and not for performing and expressing estimates. A simple joke tells the story. A museum attendant was overheard telling a group of schoolchildren that a fossil bone in the collection is sixty-five million and twelve years old. Someone asked him how he could be so precise, and he explained that when he came on the job, twelve years previously, he was told that the bone is sixty-five million years old. So he then did the sum:

```
65,000,000
     + 12
65,000,012
```

The paradox here is that although in context the sum is ridiculous, our system of arithmetical notation constrains us to write it that way. We need a simple notation in which our calculations can reflect, and not violate, our common sense; and where we can express our estimates with clarity.

The notational system "NUSAP" enables the different sorts of uncertainty in quantitative information are displayed in a standardized and self-explanatory way. It enables providers and users of such information to be clear about its uncertainties. Since the management of uncertainty is at the core of the quality-control of quantititative information, the system "NUSAP" also fosters an enhanced appreciation of the issue of quality in information. It thereby enables a more effective criticism of quantitative information by clients and users of all sorts, expert and lay.

The NUSAP system is based on five categories, which generally reflect the standard practice of the matured experimental sciences. By providing a separate box, or "field", for each aspect of the information, it enables a great flexibility in their expression. By means of NUSAP, nuances of meaning about quantities can be conveyed concisely and clearly, to a degree that is quite impossible otherwise. The name "NUSAP" is an acronym for the categories. The first is Numeral; this will usually be an ordinary number; but when appropriate it can be a more general quantity, such as the expression "a million" (which is not the same as the number lying between 999,999 and 1,000,001). Second comes Unit, which may be of the conventional sort, but which may also contain extra information, as the date at which the unit is evaluated (most commonly with money). The middle category is Spread, which generalizes from the "random error" of experiments or the "variance" of statistics. Although Spread is usually conveyed by a number (either $\pm$, % or "factor of")it is not an ordinary quantity, for its own inexactness is not of the same sort as that of measurements.

This brings us to the more qualitative side of the NUSAP expression. The next category is Assessment; this provides a place for a concise expression of the salient qualitative judgements about the information. In the case of statistical tests, this might be the significance-level; in the case of numerical estimates for policy purposes, it might be the qualifier "optimistic" or "pessimistic". In some experimental fields, information is given with two $\pm$ terms, of which the first is the spread, or random error, and the second is the "systematic error" which must estimated on the basis of the history of the measurement, and which corresponds to our Assessment. It might be thought that the "systematic error" must always be less than the "experimental error", or else the stated "error bar" would be meaningless or misleading. But the "systematic error" can be well estimated only in retrospect, and then it can give surprises. Fig. 1 shows how the successive recommended values of a well-known fundamental physical constant bounced up and down, by rather more than the "experimental error", several times before settling down.
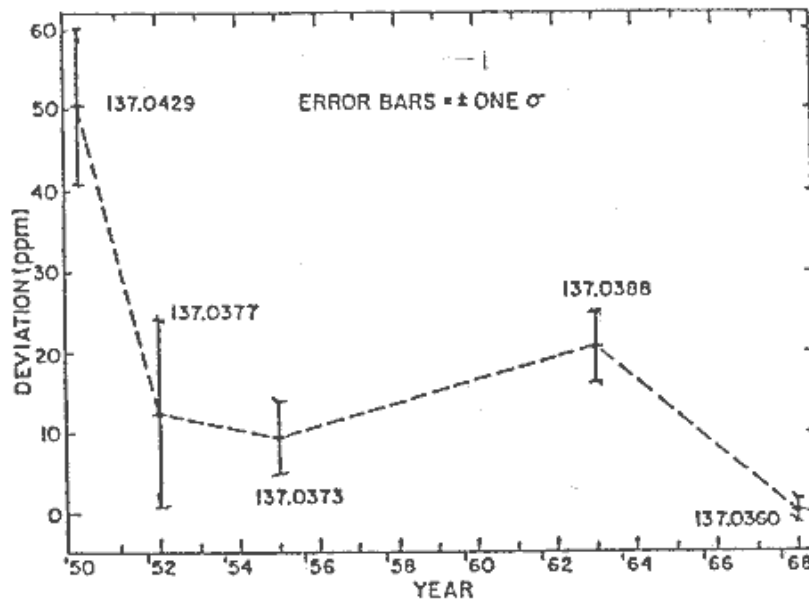
Fig. 1. Succesive "recommended values" of the fine-structure constant.

Finally there is P for Pedigree. It might be surprising to imagine numbers as having pedigrees, as if they were showdogs or racehorses. But where quality is crucial, a pedigree is essential. In our case, the pedigree does not show ancestry, but is an evaluative description of the mode of production (and where relevant, of anticipated use) of the information. Each special sort of information has its own pedigree; and we have found that research workers can quickly learn to formulate the distinctions around which a special pedigree is constructed. In the process they also gain clarity about the characteristic uncertainties of their own field. Although we have not yet had the opportunity to test this with non-expert groups, we are equally sure with some preliminary training, any client or user could learn how to elicit the pedigree of information being provided by an expert.

The pedigree is expressed by means of a matrix; the columns represent the various phases of production or use of the information, and within each column there are modes, normatively ranked descriptions. These can be numerically graded, so that with a coarse arithmetic, a "quality index" can be calculated for use in Assessment if desired. For general statistical information, the pedigree is laid out as in the Table, where the top row has grade 4 and the bottom two, 0. For a numerical evaluation, average scores of 4 downwards are rated as High, Good, Medium, Low and Poor.

�

Grade	Definitions & Standards	Data-collection & Analysis	Institutional Culture	Review
4	Negotiation	Task-force	Dialogue	External
3	Science	Direct Survey	Accommodation	Independent
2	Convenience	Indirect Survey	Obedience	Regular
1	Symbolism	Educated Guess	Evasion	Occasional
0	Inertia	Fiat	No-contact	None
0	Unknown	Unknown	Unknown	Unknown

The Pedigree Matrix for Statistical Information

The first column describes how the job is defined, for any competent statistician knows that "just collecting numbers" leads to nonsense. The whole Pedigree matrix is conditioned by the principle that statistical work is (unlike some traditional lab research) a highly articulated social activity. So in "Definition and Standards" we put "negotiation" as superior to "science", since those on the job will know of special features and problems of which an expert with only a general training might miss. It is important to be able to descibe low-quality work; and so "symbolism" in statistics is something which any comprehensive scheme must allow for. Similarly, a "Task Force" gets a higher rating than a "Direct Survey", for the latter (like a census) may produce information that is not tailored to the problem at hand. The other two columns relate to the more directly social aspects of the work. "Institutional Culture" describes the relations between the various levels in hierarchy of command and control; and we allow for the phenomena variously described by "Clochemerle" or "Schweik". Since quality-assurance is an essential part of any productive process, including that for information, we have a column for "Review". This needs no explanation.

Thus the pedigree matrix, with its multiplicity of categories, enables a considerable variety of evaluative descriptions to be simply scored and coded. In our book Uncertainty and Quality in Science for Policy (Kluwer Academic, 1990), we illustrated the NUSAP system with a somewhat imaginary example of the history of the statistics on hand-pumps for drinking water in a Third-World country. The earlier efforts had a distinctly low Pedigree profile; inertia, symbolism and fiat were prominent, along with the absence of effective review; but by the end, with the lessons of exprience, improvements could be recorded.

The use of NUSAP could also highlight crucial features of the process. For example, the problems of definition can be explored; is a "pump" one that is listed in an old census, one that is ordered from abroad, one that is registered as delivered, one that is installed, or one that is actually in full and satisfactory use? At the other end of the process, NUSAP alerts us to the meaningfulness or otherwise of numerical expressions. If we see a number like 11,287 coming out of an unsophisticated statistical exercise, we should be able to recognize hyper-precision. On the other hand, a correctly framed estimate like "11,300:pumps: $\pm 5\%$" has certain uses, but the less precise statement, "<111/2:K-pumps" may be more appropriate for policy purposes. In prose, this is "somewhat less than 111/2 thousand pumps, where the aggregated unit of counting is a thousand-pumps, with a Spread of a quarter-thousand. Thus with NUSAP we are able to provide numerical statements with considerable nuance of expression.

We know enough about the use of numbers to be aware that no single system will prevent incompentence and abuse of statistics. But with the improvement of competence all around, and especially with the arming users and clients with an instrument of analysis, NUSAP will at least make it possible for the debate to be conducted at a higher level. We cannot claim that an improvement in this one area of practice will transform industry and administration for the better; but we do believe that everyone will be better off when they know what they are doing, in the management of the uncertainty and quality of their quantitative information.

(Note on references: all citations will be found in the book Uncertainty and Quality in Science for Policy, by S.O. Funtowicz and J.R. Ravetz)