Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Reliability Engineering and System Safety xx (0000) xxx–xxx
RELIABILITY ENGINEERING & SYSTEM SAFETY
www.elsevier.com/locate/ress

The anatomy of the squizzel☆
The role of operational definitions in representing uncertainty

Roger Cooke

Department of Mathematics, Delft University of Technology, Delft, The Netherlands

Abstract

An anecdotal introduction of the role of operational definitions in representing uncertainty is followed by a brief history of operational definitions, with particular attention to the foundations of probability and the fuzzy representations of uncertainty. A short summary of experience at the TU Delft points to relevant open questions. These in turn are illustrated by a recent application to NO_x emissions in The Netherlands.

Keywords: Operational definitions; Rational decision theory; Fuzzy sets; Uncertainty analysis; Expert judgment; Dependence

1. Introduction

I asked Didier Dubois at a 1996 meeting of the European Fusion work group in Lecoutre, France:

> How many legs does a squizzel have?

He answered:

> First tell me what a squizzel is.

Right answer.

But instead of telling him, I said:

> Well, just use your own idea of what you think a squizzel is, and tell me how many legs it has.

The way he felt then is the way I feel when someone asks:

> What is John's fuzzy membership in the set of tall people?

or

> What is the degree of possibility that the Loch Ness monster exists?

How is fuzzy membership or degree of possibility defined? I am not asking for a mathematical definition, I am asking for an operational definition, that is, for a rule, which indicates how the mathematical notions are intended to be interpreted. Leaping over more than a century of semantic analysis, a modern rendering of this question is the following.

If Bob says:

> The fuzzy membership of John in the set of tall people is 0.7057.to what sentences in the natural language not involving the word 'fuzzy' is Bob committee?

If the set of sentences given in answer to this question is the empty set, then this is operationally equivalent to the anatomy of the squizzel.

In this paper, I give a crash course in the modern philosophy of science, reviewing operational definitions, their role in the representation of uncertainty and in 'degenerating problem shifts'. I then draw some lessons from the over 10,000 expert elicitations performed according to the probabilistic method of the TU Delft. I explain why, in my view, the challenge problems miss the challenges confronting the representation of uncertainty today.

2. Operational definitions: history

It impressed me that most members of the FUSION workgroup, and many advocates of 'alternative' representations of uncertainty seem unfamiliar with the notion

☆ This paper is based on a Discussion Paper for FUSION workgroup, 1996.

E-mail address: r.m.cooke@its.tudlft.nl (R. Cooke).

0951-8320/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.ress.2004.03.019

RESS 3348—25/3/2004—12:24—SHYLAJA—98663— MODEL 5

of operational definitions. Here is a very brief historical sketch.

The notion can be traced back at least to the first rational reconstructions of classical mechanics by Mach [20] and Hertz [12]. Both authors were troubled by conceptual problems associated with the notions of force, absolute time and space. Mach's approach was to 'deconstruct' the mechanics of Newton by semantic analysis. Defining 'meaning' as correspondence with sensations, he tried to rebuild the theory using only primitive terms which could be directly associated with sensations. His efforts led him to the conclusion that the notions of absolute time and space had no meaning and could be dropped.

Einstein's 1905 article on special relativity explicitly applies Mach's semantic analysis to deconstruct the classical notion of simultaneity and thus unified classical mechanics and classical electromagnetism in special relativity. Semantic analysis has been applied many times to eliminate redundant or meaningless concepts, which block progress. After Einstein, the second most spectacular example is Niels Bohr's resolution of the wave-particle duality.

Hertz took a very different approach. He axiomatized classical mechanics as a formal system and provided explicit rules for interpreting the primitive terms of this system in terms of measurements. Hertz' mechanics, via the early work of Ludwig Wittgenstein, was the source of 'formal philosophy of science'. Within this discipline, one axiomatizes theories in order to study their properties and to clarify their interpretation. This has evolved into a picture of theories as layers of two (or more) languages. At the lowest level is an observation language in which the results of elementary measurements are described (e.g. 'the needle points to 4'). A 'theoretical language' with axioms may contain terms which are directly interpreted in the observational language, but it's terms may also be interpreted in more complicated ways. However, to be non-trivial, a theory must entail observational statements which can be checked by experiment.

The term 'operational definition' was actually introduced by Bridgeman in 1927, and similar concepts may be found in the writings of many philosophers of science ('coordinating definitions', 'semantic rules', 'correspondence rules', 'epistemic correlations', and 'rules of interpretation' (see Ref. [21]). Bridgeman's concept is somewhat naive in so far as it recognizes only the simplest way of giving meaning to abstract expressions. However, his term has perhaps the widest currency and seems suitable (to me) for discussing the meaning of probability, possibility, fuzziness and related concepts.

It has long been recognized that there is no 'theory free' or 'presupposition-less' interpretation of languages. Rather, we interpret a theory, expressed in a theoretical language, in terms of another language, perhaps even the 'ordinary language'. The point of doing this is to link up with wider linguistic communities—that is more or less the history of semantic analysis in a nutshell.

3. Operational definitions: probability

The above suggests that giving operational definitions is an essential part of the foundations for a discipline. The period from 1900 to say 1940 was marked by intense activity at the foundations of probability. It will be clear that the choice of axioms is intimately related to the choice of interpretation. Axioms are evaluated not only with regard to consistency, but also with regard to their ability to describe accurately the intended interpretation. Four main types of interpretations of probability have been proposed.

3.1. Classical interpretation

This is generally attributed to Laplace [19] who defined probability as, 'the number of favorable cases divided by the number of equi-possible cases'. Examples from coin tossing and dice-throwing were used to illustrate what is meant by 'equi-possible'. The fact that we no longer hear about this interpretation is related to the inability of its proponents to provide an operational definition of equi-possible. Each proposed operational definition was met by counterexamples and paradoxes.

3.2. Logical interpretation

This was first proposed by Keynes [15] and later taken up by Carnap [4,5]. The idea was that conditional probability should be interpreted as partial entailment. The notion of partial entailment never received a satisfactory interpretation and this interpretation is generally regarded as dead.

3.3. Frequentist interpretation

Von Mises [29] advanced the interpretation of probability as limiting relative frequencies in a 'collective' or 'random sequence'. The reference to collectives or random sequences is essential. For example, the relative frequency of '1's' in the sequence:

$$1, 0, 1, 0, 1, 0, 1, 0, \ldots$$

would not be interpreted as a probability. The frequency interpretation in fact introduces 'probability' as a defined notion in a new formal system with a new primitive term collective. Of course, he is obliged to give an operational definition of collective. Although he could point to 'random looking' sequences, a good axiomatization with operational definitions was never proposed. Later Kolmogoroff, Martin-Lof, Schnorr [28] and others did succeed in this. Very roughly, a random sequence is one which passes all 'recursive statistical tests'.

3.4. Subjective interpretation

The above three interpretations, or perhaps we should say 'interpretation programs' are 'objective' in the sense that the probability assigned to an event should hold for all individuals…it should be a matter of rational consensus.

The subjective interpretation interprets probability in terms of degree of belief of a subject. Different subjects can have different degrees of belief for one and the same event.

Borel [2] and Ramsey [25] are regarded as founders of this interpretation. The best exposition is by Savage [27]. Probability is interpreted as 'degree of belief of a rational subject'. This degree of belief is measured by observing an individual's choice behavior in specific situations. I will recall a few points here.

There are many ways to operationalize 'degree of belief', but I believe that Savage's is the best from a philosophical viewpoint. Degree of belief is interpreted in terms of rational preference, and preference is operationalized in terms of choice behavior.

Consider two events:

F : France wins the next World Cup Soccer tournament
U : The USA wins the next World Cup Soccer tournament.

Consider two lottery tickets:

LF: worth $10,000 if F, worth $100 otherwise
LU: worth $10,000 if U, worth $100 otherwise.

John is offered ONE of these, and he may choose whichever he wants. Now,

John's degree of belief in F at least as great as his degree of belief in $U''$is operationalized as

John chooses LF in the above choice situation

We will denote 'John chooses LF in the above choice situation' as LF ≧ .LU (this is Savage's notation).

The following can be proved:

(i) If John's preferences satisfy the 'principle of definition' (my term for one of Savage's axioms), then the degree of belief does not depend on the values used in the lottery (we can use $30,000 instead of $10,000; in fact we do not need money at all, we can use any pair of consequences, as long as one is 'better than' the other).

(ii) If John's preferences satisfy the dominance axiom, then his degree of belief in any event is less than or equal to his degree of belief in the trivial event, and is greater than or equal to his degree of belief in the empty event.

(iii) If John's preferences satisfy the sure thing principle, then his degree of belief is additive in the following sense: if $F \cap B = U \cap B = \emptyset$, then

$$\text{Deg Bel}(F) \geq \text{Deg Bel}(U) \text{ if and only if}$$

$$\text{Deg Bel}(F \cup B) \geq \text{Deg Bel}(U \cup B).$$

(iv) If John's preferences are transitive and satisfy a technical axiom, then there exists one and only one finitely additive probability measure P which represents John's degree of belief in this sense: for all events A, B :

$$\text{Deg Bel}(A) \geq \text{Deg Bel}(B) \text{ if and only if } P(A) \geq P(B)$$

The reader should test these axioms against his/her own preference structure. Take for this purpose:

B : Belgium wins the next World Cup Soccer tournament

In particular, the reader should verify whether for him/her:

LF ≧ .LU and LU ≧ .LB imply LF ≧ .LB

LF ≧ .LU implies $L(F \cup B) \geqq .L(U \cup B)$.

If so, then modulo a technical axiom, the reader's uncertainty is represented by a unique (subjective) probability measure.

The axioms mentioned above characterize what Savage means by rational preference. Every axiom has been discussed, tightened, relaxed, etc. and numerous variations of this theory have been explored. Interesting as these are, they remain variations on a theme, and the theme is the representation of degree of belief as a finitely additive probability.

This does not mean that there are no problems with this theory. No formal theory will be wholly adequate to an informal concept, Cooke [6] elaborates some aspects of partial belief which are not captured by subjective probability.

4. Operational definitions: possibility and fuzziness

Given the enormous intellectual effort put into operationalizing the notion of probability, someone from my background is unable to understand why the proponents of alternative representations of uncertainty show so little interest in operational definitions. What does it mean to say:

(I) The possibility that France wins the next World Cup Soccer tournament is greater than the possibility that Belgium wins the next World Cup Soccer tournament or

(II) France belongs to the set of winners of the next World Cup Soccer tournament with fuzzy membership 0.7, and Belgium belongs to the set of winners of the next World Cup Soccer tournament with fuzzy membership 0.6

Why is it not a simple matter to say what these mean? At least, it would be nice to have answers to the following questions:

(1) Are possibility and fuzzy membership objective or subjective?
(2) Is there any implication between statements (I) or (II) (perhaps relative to a given individual) and (the individual's) choice behavior, e.g. with the lotteries LB and LC?
(3) Is there any implication between (I) or (II) and any other statements in a language not containing the words 'possibility' and 'degree of membership' or their synonyms?

In the absence of operational definitions, we can only evaluate the suitability of such putative representations of uncertainty on the basis of formal properties. Many such representations, including fuzziness and possibility, have a feature called 'truth functionality' which render them highly unsuitable as generally applicable representations of uncertainty. Truth functionality says that the uncertainty in proposition *A* AND *B* is some function of (only) the uncertainty of *A* and the uncertainty of *B*, and similarly for *A* OR *B*. Thus, for example[1] Ayyub [1] proposes a fuzzy representation of the degree of belief or 'membership uncertainty' that element *x* belongs to set *A*; the membership uncertainty that *x* belongs to *A* AND *B* is the minimum of the membership uncertainties that *x* belongs to *A* and that *x* belongs to *B*.

To appreciate what this means, consider the following example. I get an email from an unknown 'Quincy'. My degree of belief or membership uncertainty that Quincy belongs to the set 'MEN' is 1/2, and my degree of belief that Quincy is a WOMEN is also 1/2. Therefore, my degree of belief that Quincy is a MAN AND WOMAN is the minimum of 1/2 and 1/2, or 1/2.

Let me emphasize that I do *not* claim that operational definitions of degree of possibility or fuzzy membership are impossible. Dubois [10] makes a very significant attempt to provide a Savage-style foundation for possibility theory. Quoting from this reference: 'By providing an act-driven axiomatization of possibility and necessity measures, possibility theory…becomes an observable assumption that can be checked for the actual behavior of a decision-maker choosing among acts, just like subjective probability after Savage axiomatics. This is why the result of this paper is significant from the view of Artificial Intelligence as laying some foundations for qualitative decision theory' (p. 477).

The value of Dubois' operational definition is that it allows us to see what a representation of uncertainty as degree of possibility or fuzzy membership[2] means in terms of preference. Roughly, it means that the preference between actions is determined by their best (or worst) consequence. Thus, the main axiom behind the possibility representation of uncertainty entails the following

> If you prefer \$100 to a fair coin bet: $B = [+\$1{,}000{,}000$ if heads, $-\$1{,}000{,}000$ if tails], and if you prefer \$100 to \$99, then you should prefer \$100 to [+\$1,000,000 if heads, \$99 if tails].

It is not impossible that such preference behavior is empirically or normatively valid in some contexts,[3] and in such contexts a possibilistic or fuzzy representation of uncertainty might be appropriate.

5. Progressive versus degenerating problem shifts

According to the Methodology of Scientific Research Programs of Lakatos [17], scientific theories do not simply get verified or falsified by experiment. Rather, research programs compete with each other in more complex ways. A program experiences a progressive problem shift when it generates new concepts, predicts new phenomena, unifies diverse fields, etc. It goes into a degenerating problem shift when it is forced increasingly to react to new developments from other programs, and is increasingly occupied with translating developments from other programs into its own terms.

This is perhaps the most realistic approach to the 'growth of knowledge'. I mention it briefly here because much of the literature surrounding alternative representations of uncertainty consists exactly in translating results and techniques from the probabilistic approach into some alternative framework. This in itself constitutes a degenerating problem shift, if not combined with the development of new ideas, methods and predictions. Up to now the ideas have been flowing *from* probability *to* alternative representations. If these are to experience a progressive problem shift, then the flow of ideas must reverse. Without operational definitions for their primitive terms, however, that will remain impossible.

---

[1] The following is taken from a review of Ayyub [1] appearing in Fuzzy Sets and Systems.

[2] Dubois considers a possibility measure as equivalent to a fuzzy membership function.

[3] The axioms of Savage's rational preference and of subjective probability are claimed to be normatively valid but not empirically valid; the describe how a rational agent *should* act and reason under un certainty, not how they actually do. In this sense, they are comparable to deductive logic.

6. Uncertainty analysis

The goal of an uncertainty analysis is a quantitative representation of uncertainty. According to the subjective interpretation, uncertainty is represented by subjective probability. Other representations may be possible, but these should be supplied with a suitable foundation, including plausible axioms and operational definitions. Although many alternative representations have been proposed, none to date have been given a foundation in this sense. For practical work, there is at present no viable alternative to the representation of uncertainty as subjective probability.

A key question is what role experts play in providing a quantitative representation of uncertainty. The method employed at the TU Delft [7] proceeds from the assumption that individual experts quantify their uncertainty on input variables, and these uncertainty distributions are combined using the 'classical model'. Expert performance is scored on 'seed' or 'calibration' variables in terms of calibration and information, and performance based weights are used to form a weighted combination of experts' distributions. The weights satisfy a proper scoring rule constraint. The analyst does not play an active role, (s)he does not alter experts' numbers and does not choose weights. Rather, (s)he merely scores the experts' assessments and combines their distributions according to pre-defined objectively traceable rules.

Other approaches are also applied. Thus, in NUREG/CR-6372 [22], it is argued that the combined expert distributions should represent the dispersion of expert views, and that the analyst should play an active role in subjectively weighting the experts to achieve this goal.

Thus, the question is who signs off on the uncertainty assessments, is it the experts or the analyst? Whoever signs off, using subjective assessments of uncertainty without verifying performance is, in my opinion, foolhardy.

7. Challenge problems

To date, there have been over 10,000 expert elicitations performed at or in collaboration with the TU Delft. These involve 29 different expert panels covering a wide variety of subjects including:

- Crane risk
- Propulsion of rockets
- Space debris
- Composite materials
- Groundwater transport
- Atmospheric dispersion and deposition
- Toxic materials
- Underground pipelines
- Transport of radiation in the soil/plants/animals
- Health/economic effects of radiation
- Failure of moveable water barriers
- Dike ring reliability
- Safety factors for airline pilots
- Montserrat eruption prediction
- Serviceability limit states
- NO_x emissions.

(For details and references, see Refs. [7,11,14]). All of these studies used seed variables: expert performance was measured in terms of calibration and information, and the performance of various combination schemes was examined.[4]

In none of these elicitations were we confronted with the bizarre situations described in the 'challenge problems' proposed for this workshop [23]. On the other hand, the problems met with in practice are not recognizable in these challenge problems. As agreed with the organizers, I will therefore extract a few lessons learned from my experience with expert judgment which might be of interest for this workshop, and sketch what I see as important problems.

8. TU Delft experience

1. Experts do not mind performance measurement.[5] It has happened repeatedly that experts defended the measurement of performance when the problem owners became nervous about publishing results.
2. Experts are leery of 'non-objective' or psychologically based methods, and are suspicious of the 'academic sandbox'.
3. Experts have no problem understanding (subjective) probability and no problem quantifying degree of belief in terms of quantiles of a subjective probability distribution. For experts without formal education (e.g. crane operators) this may require intuitive explanation.
4. Experts are *not* uniformly overconfident, though overconfidence certainly does arise. As a general rule, the more a field is based on physical measurements, the better the experts' performance and the more the experts agree.
5. It is *always* possible to find suitable calibration variables. If there are no measurements relevant to a given field, then this falls outside empirical science and outside expert judgment. The existence of god is not an issue to be adjudicated by expert judgment.
6. In general, though not always, the performance based combination of expert judgments performs better, in terms of calibration and information, than an equal weight combination and also better than the best expert (see Ref. [11]).

[4] This has not yet been done for the NO_x study [9].

[5] In our total experience, there have been may be three experts who objected to performance measurement.

Finally, I would add that experts are generally cooperative and will try to conform to the elicitation format suggested by the analyst. I suggest that reports of experts' 'inability to give distributions' reflects the attitude of the analyst, not the experts.

9. Conclusions: open issues

I conclude by identifying several important open issues involving expert judgment.

9.1. Dependence between elicition variables

We have developed techniques for eliciting dependences based on rank correlation coefficients, and applied these with some success in the USNRC-EU joint uncertainty analysis of accident consequence codes for nuclear power plants [3,8]. This is certainly not the last word on this subject.

The importance of this issue can be illustrated with the recent NO_x study [9]. Fig. 1 shows the uncertainty in emissions (kg/yr).

Suppose X_i is the emission the ith auto, where X_i is normally distributed with mean 37 and standard deviation 50 (these values are representative for Benzene Without Regulated Catalysor, though the distribution in Fig. 1 is not normal). If the uncertainty for the autos were 'aleatory' then each auto's emissions would be drawn independently from this distribution. The uncertainty in the total emissions of 2,040,000 benzene autos in the Netherlands without catalysor would be the 2,040,000-fold convolution of the distribution in Fig. 1; it would be normal with mean 75 million and standard deviation 71,414. Hence the 5% 95% uncertainty band for the total emission would be very narrow indeed: 74.9 and 75.1 million. The explanation is simple, the mean of the sum $\sum_{i=1\ldots N} X_i$ increases with N, whereas the standard deviation of $\sum_{i=1\ldots N} X_i$ increases with $\sqrt{N}$. If the uncertainty were 'epistemic', then the uncertainties would be completely correlated. The mean for 2,040,000 autos would still be 75 million, but the 90% confidence band would be 11.4–137.1 million.

Clearly neither of these alternatives is realistic. The approach in Ref. [9] cannot be explained here. Suffice to say it involved 'majorizing' the correlation between individual auto's, and resulted in a 90% confidence band shown in Fig. 2. The rank correlation between the emissions of two randomly sampled benzene auto's without catalysor was 0.04; for the emissions from two groups of 100 such auto's it was 0.57, and for two groups of 10,000 such autos it was 0.99. This sort of behavior holds quite generally. In fact, one can show that if the uncertainty in variable X_i is the sum of aleatory and epistemic random variables, then the variance of $\sum_{i=1\ldots N} X_i$ is dominated by the variance of the epistemic component *and* the covariance between the aleatory and epistemic components as N gets large. In general, aleatory and epistemic components *will* be correlated (for details, see

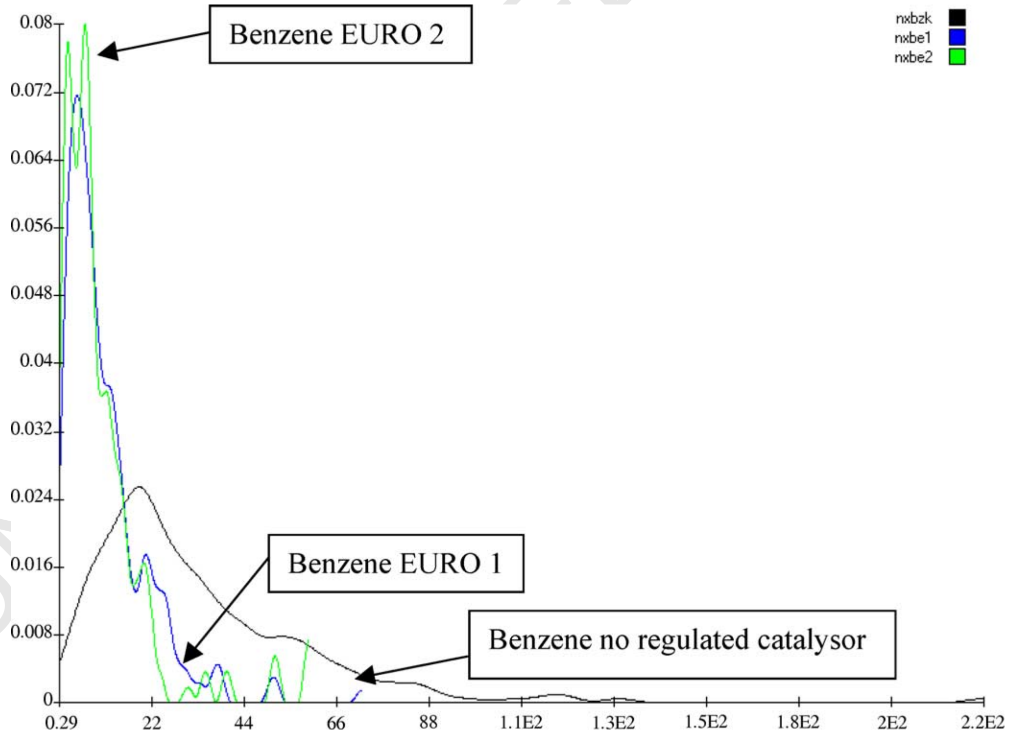

Fig. 1. Probability density function for one random auto of NO_x (kg/yr) for milieu classes: no regulated catalysor, EURO1 and EURO2.

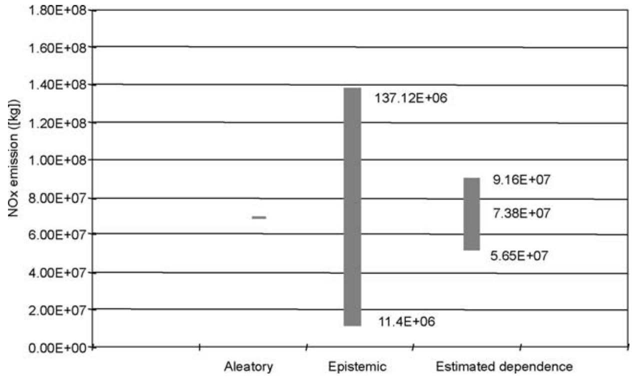

Fig. 2. Epistemic versus aleatory uncertainty.

Ref. [9]). The point is that the issue of 'aleatory versus epistemic' is really an issue of dependence. Simple solutions will only work for simple problems.

9.2. Does combination commute with model computations?

In other words, should we combine experts and then propagate this combined distribution through a model, or should we first propagate each expert's distributions through the model, and then combine. It can make a significant difference.

Fig. 3 shows the results of propagating the (three) experts' distributions separately through the emissions model, and also the results of first combining the experts (DM before) before propagating, and combining the experts distributions after propagating them through the model (DM after). The combination is with equal weights. The differences between the two DMs are significant, though within the spread among the experts themselves.

9.3. Dependence between experts

Some recent work in this area is found in Ref. [13]. Research is needed both in ways of measuring dependence, and ways of using this information in combining experts' assessments. To date, there is no satisfactory proposal.

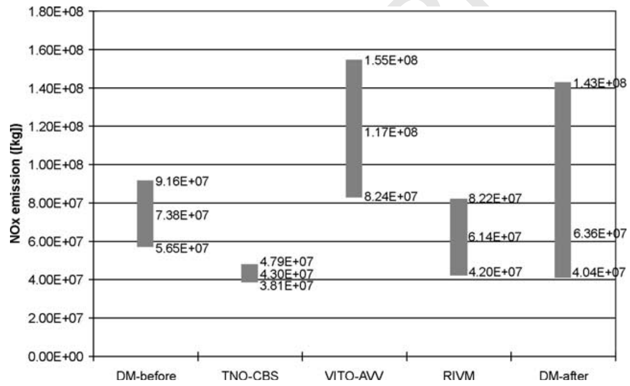

Fig. 3. Experts' propagated distributions, combining before and after propagation, using equal weights. The numbers shown are the 5, 50 and 95% quantiles.

9.4. Expert judgment for models

The question is, how can expert judgment be used to build models. In the particular case of graphical models such as Bayesian Belief Nets, how do we effectively elicit conditional independence relations from experts? Given the importance of Bayesian Belief Nets in artificial intelligence and engineering applications, it is imperative to find ways to find ways of controlling their staggering complexity and quantifying them with expert judgment in a traceable way.

10. Uncited references

[16]. [18]. [24]. [26].

References

[1] Ayyub BM. Elicitation of expert opinions for uncertainty and risks. Boca Raton, FL: CRC Press; 2001.

[2] Borel E. Elements de la theorie des probabilities, Miche. 1950.

[3] Brown J, Goossens LHJ, Harper FT, Haskin EH, Kraan BCP, Abbott ML, Cooke RM, Young ML, Jones JA, Hora SC, Rood A. Probabilistic accident consequence uncertainty analysis: food chain uncertainty assessment. Prepared for US Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6523, EUR 16771, Washington/USA, and Brussels-Luxembourg, vols. 1 and 2.; 1997.

[4] Carnap R. Der logische aufbau der welt, Berlin. 1922.

[5] Carnap R. Abriss der logistik, Wien. 1929.

[6] Cooke RM. Conceptual fallacies in subjective probability. Topoi 1986;5:21–7.

[7] Cooke RM. Experts in uncertainty: opinion and subjective probability in science. Oxford: Oxford University Press; 1991.

[8] Cooke RM, Goossens LJH. Procedures guide for structured expert judgment. European Commission, EUR 18820 EN; 2000.

[9] Cooke RM, Kraan B. Uncertainty analysis for automobile emissions NO_x: overview of methods and results, TU Delft. 2002.

[10] Dubois D, Prade H, Sabbadin R. Decision-theoretic foundations of qualitative possibility theory. Eur J Oper Res 2001;128: 459–78.

[11] Goossens LHJ, Cooke RM. Evaluation of weighting schemes for expert judgment studies. In: Mosleh A, Bari R, editors. Probabilistic safety assessment and management, vol. 3. New York: Srpinger; 1998. p. 2113–8.

[12] Hertz H. The principles of mechanics presented in a new form, New York. 1956.

[13] Jouini MN, Clemen RT. Copula models for aggregating expert opinions. J Oper Res 1996;44(3):444–57.

[14] Kallen MJ, Cooke RM. Aggregating expert opinion, TU Delft. 2002.

[15] Keynes J. Treatise on probability, London. 1973.

[16] Kuhn T. The structure of scientific revolutions. 1970.

[17] Lakatos I. History of science and its rational reconstruction. In: Buck, Cohen, editors. Boston studies in the philosophy of science VIII, Dordrecht. 1979. p. 137–45.

[18] Lakatos I. The methodology of scientific research programs, Cambridge. 1978.

[19] Laplace S. A philosophical essay on probabilities, New York. 1951.

[20] Mach E. The science of mechanics, La Salle. 1960.

[21] Nagel E. The structure of science, London. 1961.

[22] US Nuclear Regulatory Commission, Guidance on uncertainty and use of experts. NUREG/CR-6372; 1997.
[23] Oberkampf WL, Helton JC, Joslyn CA, Wojtkiewicz SF, Ferson S. Challenge problems: uncertainty in system response given uncertain parameters. Reliab Engng Syst Saf; this issue.
[24] Popper K. The logic of scientific discovery, London. 1974.
[25] Ramsey F. Truth and probability. In: Foundations of mathematics and other logical essays, London; 1931.
[26] Reichenbach H. Wahrscheinlichkeitslere, Leiden. 1935.
[27] Savage L. Foundations of statistics, New York. 1956.
[28] Schnorr C. Zufalligheit und wahrscheinlichkeit, Berlin. 1970.
[29] Von Mises R, Wahrscheinlichkeit, statistik und wahrheit, Wien; 1936.