Risk Analysis, Vol. 19, No. 1, 1999

The Expected Value of Information and the Probability of Surprise

James K. Hammitt[1] and Alexander I. Shlyakhter[2]

Risk assessors attempting to use probabilistic approaches to describe uncertainty often find themselves in a data-sparse situation: available data are only partially relevant to the parameter of interest, so one needs to adjust empirical distributions, use explicit judgmental distributions, or collect new data. In determining whether or not to collect additional data, whether by measurement or by elicitation of experts, it is useful to consider the expected value of the additional information. The expected value of information depends on the prior distribution used to represent current information; if the prior distribution is too narrow, in many risk-analytic cases the calculated expected value of information will be biased downward. The well-documented tendency toward overconfidence, including the neglect of potential surprise, suggests this bias may be substantial. We examine the expected value of information, including the role of surprise, test for bias in estimating the expected value of information, and suggest procedures to guard against overconfidence and underestimation of the expected value of information when developing prior distributions and when combining distributions obtained from multiple experts. The methods are illustrated with applications to potential carcinogens in food, commercial energy demand, and global climate change.

KEY WORDS: Probability; uncertainty; data; risk assessment.

1. INTRODUCTION

Risk assessors attempting to use probabilistic approaches to describe uncertainty often find themselves in a data-sparse situation: available data are only partially relevant to the parameter of interest, so one needs to adjust empirical distributions, use explicit judgmental distributions, or collect new data. In determining whether or not to collect additional data, whether by measurement or elicitation of experts, it is useful to consider the expected value of the information (Raiffa, 1968; National Risk Council, 1996; Presidential/Congressional Commission on Risk Assessment and Risk Management, 1997). The expected value of information depends on the prior distribution used to represent current information. If the prior distribution is too narrow, the calculated expected value of information will be biased; in cases of interest to risk analysts, the bias is likely to be downward. The well-documented tendency toward overconfidence (Kahneman *et al.,* 1982; Morgan and Henrion, 1990), including the neglect of potential surprise, suggests this bias may be important. This paper examines the expected value of information and suggests procedures to guard against overconfidence and underestimation of the expected value of information.

The word ''uncertainty'' is often used without formal definition. Unless otherwise noted, the term will be used here in the sense defined by Rothschild and Stiglitz (1970). Their concept applies to distributions with equal expected values and can be expressed

[1] Department of Health Policy and Management and Center for Risk Analysis, School of Public Health, Harvard University, 718 Huntington Ave., Boston, Massachusetts 02115; e-mail: James_Hammitt@Harvard.edu

[2] Department of Physics, Harvard University, Cambridge, Massachusetts 02138; e-mail: shlyakhter@physics.harvard.edu

0272-4332/99/0200–0135$16.00/1

using any of three mutually equivalent definitions: a random variable y is more uncertain than another random variable z if: (a) y is equal to z plus random noise, (b) every risk averter prefers a gamble with payoffs equal to z to one with payoffs equal to y, or (c) the density of y can be obtained from the density of z by shifting weight to the tails through a series of mean-preserving spreads. The Rothschild-Stiglitz definition is equivalent to the ordering imposed by second-order stochastic dominance (e.g., Huang and Litzenberger, 1988). The Rothschild-Stiglitz definition provides only a partial ordering over distributions; it is possible that neither one of a pair of distributions expresses more uncertainty. Moreover, although a distribution that represents greater uncertainty has a larger variance, a distribution with a larger variance need not represent greater uncertainty.

The paper is organized as follows. Empirical evidence of overconfidence in probability assessment and alternative distributional forms that better incorporate the probability of surprise are reviewed in the Sect. 2. Section 3 examines some effects of the prior distribution on the expected value of information (EVI), including the tendency for overconfidence to bias calculations of EVI downward, and illustrates this effect in the context of commercial energy projections. Section 4 describes four heuristic factors that influence the expected value of information and illustrates their effect with an example of a potentially carcinogenic food additive. In Sect. 5, alternative methods for combining distributions representing expert judgment or other data are reviewed and their results compared using expert judgment about global climate change. Conclusions are presented in Sect. 6.

2. EVIDENCE AND MODELING OF OVERCONFIDENCE

The importance of incorporating surprise into uncertainty modeling and analysis stems from the well-documented tendency of both experts and lay people to underestimate uncertainty in their knowledge of quantitative information. When asked to construct prior distributions for quantities that can be verified, subject matter experts and lay people often produce distributions that are far too tight. Typical results find 55–75% of true values outside subjective interquartile ranges and 20–45% outside central 98% confidence regions (Alpert and Raiffa, 1982; Lichtenstein *et al.,* 1982; Morgan and Henrion, 1990).

Henrion and Fischhoff (1986) examined the history of measurements of fundamental physical constants (e.g., the speed of light). New and more precise measurements are often outside the reported error bars of the old measurements. To quantify the extent to which subsequent measurements depart from the distributions implied by previous measurements, Henrion and Fischhoff introduced the Surprise Index (SI), defined as the fraction of new results falling more than 2.3 times the (old) reported standard deviation from the old result. If new measurements were randomly drawn from a normal distribution with mean and standard deviation equal to the old values, the SI would be 2%. Analysis of physical measurements suggests the SI varies between about 10 and 40% (Henrion and Fischhoff, 1986).

In some fields, experts have been shown to provide relatively well-calibrated probability judgments. The classic example is meteorology, where forecasts of precipitation probabilities and of maximum and minimum daily temperatures have been shown to be well calibrated (Murphy and Winkler, 1977). Meteorologists benefit from experience in forecasting probabilities for a large number of similar events and receiving rapid feedback. More recently, Winkler and Poses (1993) have shown that physicians' estimates of patient survival probabilities are reasonably well-calibrated, even though physicians do not routinely provide quantitative survival probabilities. In contrast, financial analysts have been shown to significantly overestimate corporate earnings growth (Chatfield *et al.,* 1989; Dechow and Sloan, 1997). In one of the few studies evaluating calibration in risk analysis, Hawkins and Evans (1989) found that industrial hygenists provided reasonably accurate estimates of the mean and 90th percentile of a distribution of personal exposure to chemical industry workers, although they substantially overestimated median exposure.

Shlyakhter (1994) and Shlyakhter *et al.* (1994) analyzed several large historical datasets of physical measurements, population, and energy projections. They concluded that empirical distributions of measurement and forecast errors have much longer tails than can be described by a normal distribution and proposed a "compound" distribution that provides a more accurate description. Define the normalized error x as the difference between the new (or true) value and the old (or forecast) value scaled by the reported standard deviation of the old value (i.e., $x = (y_1 - y_0)/\Delta'$, where y_1 and y_0 are the new (realized) and old (forecast) values, respectively, and Δ' is the reported standard deviation of the old value). Assume x is normally distributed with standard deviation Δ, but that the standard deviation is incorrectly assessed as Δ' ($\leq \Delta$) because of overconfidence. Denote the extent of overconfidence by $t = \Delta/\Delta'$, and model it as randomly distributed with density function $f(t)$. The density function for x is then

$$p(x) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{f(t)}{t} e^{-(x^2/2t^2)}\, dt. \tag{1}$$

Shlyakhter (1994) proposed modeling $f(t)$ using a half-normal distribution with mode equal to one,

$$\begin{aligned} f(t) &= \sqrt{\frac{2}{\pi}} \frac{1}{u} e^{-[(t-1)^2/2u^2]}, \qquad t > 1 \\ f(t) &= 0 \qquad \text{otherwise.} \end{aligned} \tag{2}$$

Values of $t < 1$ represent underconfidence and are assigned zero probability. Substituting Eq. (2) into (1) and integrating yields the cumulative probability of deviations exceeding $|x|$

$$S(x) = \sqrt{\frac{2}{\pi}} \frac{1}{u} \int_1^\infty e^{-[(t-1)^2/2u^2]}\, erfc\left(\frac{|x|}{t\sqrt{2}}\right) dt$$

where *erfc* is the complementary error function

$$erfc(z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-\xi^2}\, d\xi.$$

The parameter u is a measure of the uncertainty of the ratio t of "true" to assessed standard deviation (Eq. 2). It may be loosely interpreted as the ratio of unsuspected errors to those that are accounted for. Note that for $u = 0$, $f(t)$ is a degenerate distribution with probability mass one assigned to $t = 1$ and the compound distribution (Eq. 1) reduces to the normal distribution.

As illustrated in Fig. 1, the tails of the compound distribution are much heavier than those of the normal distribution. For example, the probability assigned to values more than two assessed standard deviations from the mean increases from 0.05 under the normal distribution ($u = 0$) to about 0.25 for $u = 1$ and to almost 0.5 for $u = 3$. Comparison of prior and subsequent measurements suggests that for physical and environmental measurements $u \approx 1$ and for projections of population growth and energy consumption $u \approx 3$ (Shlyakhter, 1994; Shlyakhter *et al.,* 1994).

The compound distribution provides a reasonable description of empirical error frequencies in the domains to which it has been applied. In future work, it would be valuable to examine alternative distributions to determine if better fits to empirical error distributions can be obtained, and if so, under what circumstances. Among the alternatives to consider would be alternative forms for the prior distribution of t (Eq. 2). Alternative priors might allow values of t between 0 and 1 to reflect the fact that forecasts are sometimes underconfident. Alternatively, t^2 could be modeled as following an F distribution. The F distribution describes a ratio of two χ^2 random variables (which are in turn sums of squared normal random variables); it is thus a reasonable model for a ratio of variances.

3. THE PRIOR DISTRIBUTION INFLUENCES THE EXPECTED VALUE OF INFORMATION

One danger in neglecting the possibility of surprise in developing prior distributions is that the Expected Value of Information (EVI) calculated using that prior may be biased downward (Hammitt, 1995). As shown below, the expected posterior probability of an event equals the prior probability (e.g., the probability that perfect information about the carcinogenic risk of specified exposure to a compound will reveal that the risk exceeds the 95th percentile of its prior distribution is exactly 5%).[3] Consequently, if the tails of the prior distribution are too thin to represent the "true uncertainty," the apparent probability of learning that the parameter is far from the prior median will be "too small." The EVI is the integral over all possible posterior distributions of the opportunity loss prevented by improved information, weighted by the probability of that information. If posterior distributions that lead to a change in decision are given too little probability, the EVI may be underestimated. Although there is no general relation between overconfidence and bias in EVI, in many risk-analytic problems the bias is likely to be downward.

In the following subsections, we examine the relationship between the prior distribution and EVI and compare the Expected Value of Perfect Information (EVPI) based on alternative prior distributions

[3] Note that a prior distribution for cancer potency must incorporate uncertainties about the form of the dose–response function and other factors. The standard 95th percentile confidence limit reported for carcinogenic potency reflects only the sampling variability in a bioassay and assumes that the linearized multistage model is appropriate.

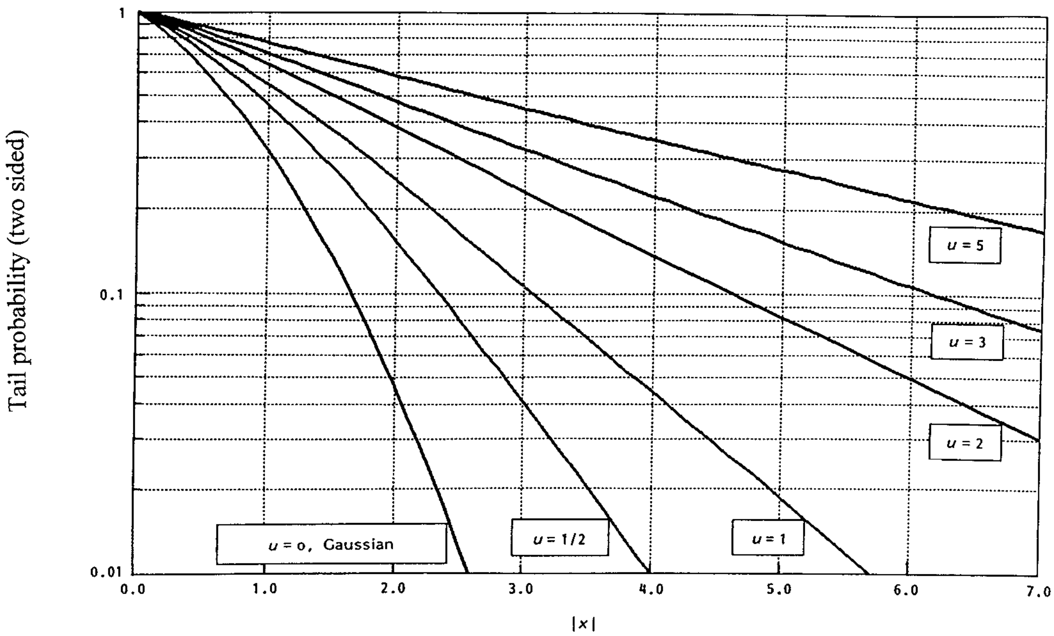

Fig. 1. Comparison of tail probabilities for compound distribution using alternative values of *u*.

to empirical estimates of EVPI for historical projections of 1990 energy consumption.

3.1. Theory

The influence of the prior distribution on the probability of reaching alternative posterior distributions was examined by Hammitt (1995). For an uncertain variable y with probability density q, let $E_q(y)$ and $V_q(y)$ denote the mean and variance, respectively. Suppose one wishes to estimate the value of a parameter $\theta \in \Theta$ where Θ is the real line or a convex subset of it. Current information about the value of θ is summarized by the prior density $f(\theta)$. An experiment is available that will estimate θ without bias but subject to measurement error. Represent the outcome of the experiment by a random variable $x \in X$, with known distribution $g(x|\theta)$ and $E_g(x|\theta) = \theta$. Once the experimental outcome x' is observed, one can derive the posterior density function $p(\theta|x')$ using Bayes' rule,

$$p(\theta|x') = \frac{f(\theta)g(x'|\theta)}{h(x')}$$

where $h(x)$ is the predictive density for x,

$$h(x) = \int_{\theta\in\Theta} g(x|\theta)f(\theta)d\theta.$$

A fundamental relationship is that the posterior density expected before the experiment is conducted is equal to the prior density:

$$\begin{aligned} E_h[p(\theta|x)] &= \int_{x\in X} p(\theta|x)h(x)dx \\ &= \int_{x\in X} \frac{f(\theta)g(x|\theta)}{h(x)} h(x)dx \qquad (3) \\ &= f(\theta)\int_{x\in X} g(x|\theta)dx \\ &= f(\theta). \end{aligned}$$

This relationship has several implications. First, whatever the likelihood function associated with a particular experiment, before the results are known the expected posterior probability content of any specified region is exactly the prior probability content of that region (e.g., the expected posterior probability that θ exceeds the 95th percentile of its prior distribution is exactly 5%). One can expect to learn *something* from the experiment, in the sense that the posterior distribution may differ from the prior, but one cannot know *what* to expect to learn. (If one knew what change in the prior distribution to expect, the prior would not adequately represent one's current information.)

Second, although the mean of the posterior distribution can be greater than, smaller than, or equal to the mean of the prior distribution (depending on the experimental outcome x), the expected value of the posterior mean is equal to the prior mean,

$$E_h[E_p(\theta|x)] = E_f(\theta).$$

Third, the expectation of the posterior variance cannot exceed the prior variance, because the prior variance is equal to the sum of the expected posterior variance and the variance of the posterior mean (Raiffa and Schlaifer, 1961),

$$V_f(\theta) = E_h[V_p(\theta|x)] + V_h[E_p(\theta|x)]. \qquad (4)$$

If the expected posterior variance $E_h[V_p(\theta|x)]$ is small (the experiment is expected to precisely estimate θ), then the variance of the posterior mean $V_h[E_p(\theta|x)]$ is large and the posterior mean is likely to differ substantially from the prior mean. In the limit, if the experiment will yield perfect information about θ, the prior distribution for the posterior mean is equal to the prior distribution for θ. Alternatively, if the experiment is not expected to reduce the variance much, the variance of the posterior mean is small and the posterior mean is not expected to differ much from the prior mean.

The expected value of information depends on the set of alternative decisions that are available and on how the payoff depends on the decision and the uncertain parameters. Let $u(d,\theta)$ denote the utility or payoff that results from choosing decision d when there is a single uncertain parameter θ, and let d_q^* denote the decision that maximizes the expected payoff when information about θ is given by the distribution $q(\theta)$. The expected value of information provided by the experiment is

$$EVI = \int_{x\in X}[\int_{\theta\in\Theta}u(d_p^*,\theta)p(\theta|x)d\theta]h(x)dx - \int_{\theta\in\Theta}u(d_f^*,\theta)f(\theta)d\theta.$$

In words, the expected value of information is the difference between the expected payoff if one observes x and then selects the optimal decision for the posterior distribution $p(\theta|x)$, and the expected payoff if the optimal decision given the prior information $f(\theta)$ is selected.

The value of information depends on both the probability that new information will lead to a different decision than would have been selected given prior information and the increased payoff that results from changing the decision. Intuitively, one might suspect that the expected value of information would be larger if the prior distribution reflects greater uncertainty (so that overconfidence would lead to an underestimate of EVI). However, because of the complexity of the dependence of EVI on the prior distribution, likelihood function, available decisions, and their payoffs, there is no general relation between uncertainty and EVI (Hilton, 1981). To understand why, consider two examples.

Example 1. An uncertain variable θ is known to be uniformly distributed with mean zero, but its range is uncertain (i.e., $\theta \sim U[-\sigma, \sigma]$ where σ is uncertain). Let $\sigma' < \sigma$ represent an overconfident estimate of σ. Two decisions (d_0 and d_1) are available with payoffs $u(d_0, \theta) = 0$ and $u(d_1, \theta) = 1 - \theta$. Thus, d_1 is preferred if $\theta < 1$; otherwise, d_0 is optimal. Assume $\sigma > 1$ (if not, d_1 is preferred regardless of the value of θ and EVPI = 0). The expected value of perfect information about θ is given by

$$\begin{aligned}\text{EVPI} &= E[\max\{u(d_0, \theta), u(d_1, \theta)\}] \\ &\quad - \max\{E[u(d_0, \theta;)], E[u(d_1, \theta)]\} \\ &= \frac{1}{2\sigma}\int_{-\sigma}^{1}(1-\theta)d\theta - \frac{1}{2\sigma}\int_{-\sigma}^{\sigma}(1-\theta)d\theta \\ &= \frac{\sigma}{4} + \frac{1}{4\sigma} - \frac{1}{2}.\end{aligned}$$

EVPI is an increasing function of σ, so EVPI is larger the more uncertain the prior distribution. If EVPI is calculated using the overconfident estimate σ', the result will underestimate the true value of information. Moreover, if $\sigma' \leq 1$ even the possibility that d_0 might be preferred will be erroneously ignored.

Example 2.[4] Maintain the same assumptions as in Example 1, but replace d_1 with d_2 where $u(d_2, \theta) = 1 - \theta^2$. For $\sigma \leq 1$, d_2 dominates d_0 and the expected value of perfect information is zero. For $1 < \sigma \leq \sqrt{3}$, EVPI = $2/(3\sigma) + \sigma^2/3 - 1$; in this region, EVPI increases geometrically with prior uncertainty and overconfidence yields an underestimate of EVPI. But for $\sigma \geq \sqrt{3}$, EVPI = $2/(3\sigma)$ and EVPI decreases with greater prior uncertainty. For $\sqrt{3} \leq \sigma' < \sigma$, overconfidence yields an *overestimate* of EVPI. Intuitively, d_0 is a more robust decision than is d_2 in that it provides optimal or near-optimal payoffs for most values of θ. The alternative decision d_2 is only preferred when the probability that $|\theta| < 1$ is larger than $1/(\sqrt{3}) \approx 0.6$. Overconfidence implies an overestimate of the probability that $|\theta| < 1$ and thus an overestimate of the probability that improved information will lead to rejecting the current choice d_0 in favor of d_2.

[4] This example was suggested by Charles Linville.

Although there is no general relation between the degree of prior uncertainty and the expected value of information, EVI is larger the greater the prior uncertainty for a class of problems that are important in risk analysis. If the payoffs to alternative decisions are linear functions of the uncertain variable (as in Example 1) then the EVI is an increasing function of the prior uncertainty[5] (Gould, 1974). In situations where the objective is to minimize the expected risk and risk under each control option is proportional to an uncertain variable, overconfidence will lead to an underestimate of EVI. In many risk-analytic problems, the expected risk is at least approximately proportional to numerous uncertain factors such as toxicity coefficients; concentrations of toxins in environmental media; inhalation, ingestion, and adsorption rates; event probabilities; and others.

3.2. Example: Energy Forecasting

To examine the effect of overconfidence on EVI, consider the problem of forecasting future energy demand. Shlyakhter *et al.* (1994) examined nearly 400 projections of 1990 energy demand by fuel type and end use. Because the true values of these variables are known, we can calculate the average realized value of information and compare it with the EVI that would have been calculated using alternative prior distributions.

EVI depends on the set of decision alternatives and the utility or loss function relating the consequences of a decision to the value of the uncertain parameter(s). For simplicity, consider the standard estimation problem where the decision set is the real line and the loss function is the squared difference between the forecast and realized values. Under these assumptions, the EVPI is the expected squared error (i.e., the variance of the prior distribution). Because we are interested in the pattern of errors rather than their absolute magnitude and wish to average across projections to estimate the mean squared error, we normalize the forecasts by their reported standard deviations and examine the variance of the normalized errors.

Forecasts for 1990 energy demand by consumption sector and fuel type were made in 1983, 1985, and 1987 and reported in the Annual Energy Outlook (U.S. Department of Energy, 1992). For each measure of energy demand, three forecast values were provided: reference, lower, and upper (denoted *R, L,* and *U,* respectively). The probability content of the forecast range is not stated, but Shlyakhter *et al.* (1994) conservatively assume that L and U correspond to one-sigma confidence limits, so the probability that the true or realized value T falls between L and U is 68% (assuming a normal distribution). To account for asymmetry in the forecast distributions, the normalized error x is defined by

$$x = \frac{T - R}{U - R}, \quad T \geq R$$

$$x = \frac{T - R}{R - L}, \quad T < R.$$

Shlyakhter *et al.* (1994) find that the frequency distribution of normalized errors is independent of both the date the forecast was issued and the quantity that was forecast. Although forecasts issued in 1987 generally predicted 1990 values more accurately than did forecasts issued in 1983, the confidence limits on the 1987 forecasts are also narrower than those of the 1983 forecasts. Overall, the normalized errors are well described using the compound distribution with $u = 3$.

We calculate the EVPI as if the decision is to accept the reference value R with loss proportional to the square of the normalized error.[6] The empirical or *ex post* estimates of EVPI are presented in Table I and compared with the EVPI that would have been calculated, *ex ante,* using triangular, normal, and compound prior distributions with $u = 1$ and $u = 3$. The empirical estimates of EVPI differ by a factor of about 6.5 across forecast date with some indication of smaller bias for the most recent forecast year. Aggregated over the 3 years in which forecasts were made, the empirical EVPI is 668 times larger than the EVPI that would have been calculated assuming the prior distribution for the normalized error was normal. Equivalently, the EVPI calculated assuming a normal prior would have underestimated the true EVPI by a factor of 668. The triangular distribution, which is attractive to some risk analysts because of its simplicity, underestimates the empirical value of EVPI by a factor of 2000 (668/0.33), three times worse than the underestimate obtained using the normal distribution. The triangular distribution yields such a large underestimate because it assigns proba-

[5] Recall that uncertainty is used in the sense defined by Rothschild and Stiglitz (1970).

[6] For asymmetric uncertainty bounds ($U - R \neq R - L$), R is not the mean and so EVPI is not equal to the variance of the distribution.

Table I. Comparison of Expected Value of Perfect Information Using Alternative Prior Distributions for Energy Forecasts[a]

Date of forecast	N	Empirical	Triangular	Normal	Compound (u = 1)	Compound (u = 3)
All forecasts						
1983	128	534	0.33	1	3.56	14.6
1985	135	1221	0.33	1	3.56	14.6
1987	119	184	0.33	1	3.56	14.6
All dates	382	668	0.33	1	3.56	14.6
"Blunders" (forecasts with $\|x\| > 10$) excluded						
1983	108	20.7	0.33	1	3.54	10.4
1985	110	12.6	0.33	1	3.54	10.4
1987	97	11.2	0.33	1	3.54	10.4
All dates	315	14.9	0.33	1	3.54	10.4

[a] EVPI is calculated as the square of the standardized error x between forecast and realization.

bility zero to normalized errors larger than one; these large errors dominate the calculated EVPI under the squared-error loss function. The EVPI calculated using the compound distribution is closer to the empirical value, although even with $u = 3$ the empirical EVPI is underestimated by a factor of almost 50 (668/14.6).

The empirically estimated EVPI is sensitive to the substantial number of forecasts for which the normalized errors are extremely large. To examine the sensitivity of EVPI to large errors (which may often be the most important), we delete the 67 "blunders" (defined as cases with $|x| > 10$). The empirical estimate of the EVPI falls from 668 to about 15, a factor of 45. Even after excluding these large errors, the EVPI calculated using any of the prior distributions underestimates the empirical value, but the estimate for the compound distribution with $u = 3$ underestimates the empirical value by only a factor of about 1.5 (averaging across all forecast dates).

In summary, the expected value of information can be very sensitive to the tails of the prior distribution and the use of an inappropriately short-tailed prior may yield a gross underestimate of the expected value of information. The limiting case of a short-tailed prior is a point estimate, for which uncertainty is not recognized and the calculated EVPI is zero.

4. HEURISTIC FACTORS EXPLAINING THE VALUE OF INFORMATION

The EVI about a parameter depends on prior knowledge about the parameter, characteristics of research opportunities that can refine the prior information, the available decision options, and the utility or loss function. Hammitt and Cave (1991) identified four heuristic factors that contribute to understanding the EVI: *uncertainty* (about the parameter value), *informativeness* (the extent to which current uncertainty may be reduced), *promise* (the probability that improved information will result in a different decision and the magnitude of the resulting gain), and *relevance* (the extent to which uncertainty about the parameter contributes to uncertainty about which decision option is preferred). These factors are introduced in the first subsection and illustrated by a numerical example in the second subsection.

4.1. Heuristic Factors

Figure 2 illustrates a hypothetical decision about whether to permit or prohibit use of a food additive. The expected social cost (e.g., number of life-years lost) is plotted against the risk θ from consuming the additive (e.g., excess cancer risk) as a function of whether the additive's use is permitted. If the additive is prohibited, K represents the expected social cost associated with use of whatever substitute would be employed.

Current information about θ is represented by the probability distribution f^0; the expected social cost if the additive is permitted is L^0. If the regulatory decision rule is to permit use of the additive if and only if the expected social cost is less than the cost associated with the substitute, the additive should be permitted. Note however that there is a substantial probability that the additive is riskier than the alternative and should be prohibited.

Assume there exists some research program that will yield one of two possible outcomes, showing ei-

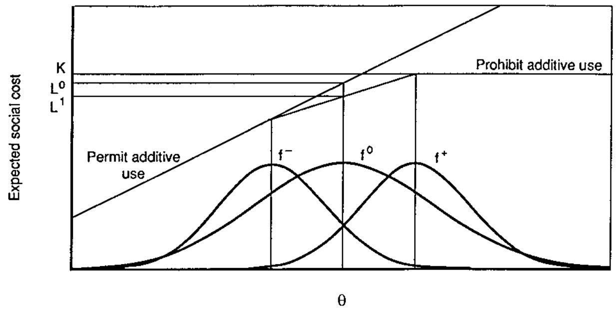

Fig. 2. The expected value of research for a decision to permit or prohibit use of a hypothetical food additive. The human health risk if the additive is permitted is θ, with current information represented by f^0. If additive use is permitted, the social cost (e.g., life-years lost) is proportional to θ if the additive is prohibited, the social loss associated with the substitute technology is K. The posterior distributions f^+ and f^- are obtained by updating f^0 with the results of a research program with two possible outcomes. The expected social cost if the additive is permitted without further research is L^0, the expected social cost if the research is undertaken and the additive is permitted or prohibited according to the result is L^1, and the EVI is L^0-L^1.

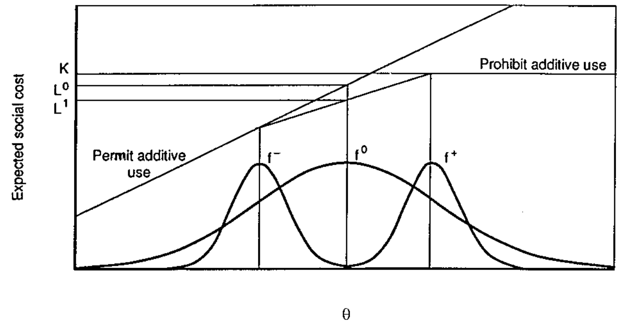

Fig. 4. Greater informativeness of the research program compared with Fig. 2 narrows the posterior distributions and spreads their means, decreasing L^1 and increasing EVI.

ther that θ is more hazardous or safer than current information suggests. The corresponding posterior distributions, f^+ and f^-, are obtained by combining the prior f^0 with the likelihood function for the research program. The expected posterior mean is equal to the mean of f^0 (by Eq. 3), the expected social cost if the research is undertaken and the additive is then permitted or prohibited as appropriate is L^1, and the EVI $= L^0 - L^1$.

The effect of greater prior uncertainty about the parameter is illustrated by Fig. 3, which represents the identical situation as Fig. 2 except the prior distribution f^0 is broader. The posterior distributions f^+ and f^-, obtained by combining the broader prior distribution with the same likelihood function as in Fig. 2, are consequently broader and have more widely separated means than in Fig. 2. As a result, the reduction in social loss conditional on learning the additive is safer than currently indicated ($L^0 - L^-$) is larger than in Fig. 2 and the EVI, $L^0 - L^1$, is also larger.

The effect of greater informativeness is shown in Fig. 4, which combines the same prior distribution as in Fig. 2 with a research program providing greater information about θ (i.e., the likelihood function conditional on each research outcome is more concentrated). Consistent with Eq. (4), the means of the posterior distributions are farther apart than in Fig. 2, and so the EVI is larger.

The effect of promise is illustrated in Fig. 5. This figure is identical to Fig. 2 except the expected social cost associated with the substitute technology (the cost incurred if the additive is prohibited) is larger than in Fig. 2. Consequently, the gain from prohibiting additive use if the research shows that it is more

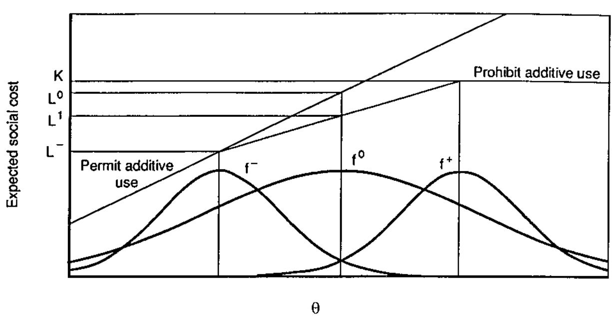

Fig. 3. Greater prior uncertainty compared with Fig. 2 spreads the means of the posterior distributions and so decreases L^1 and increases EVI.

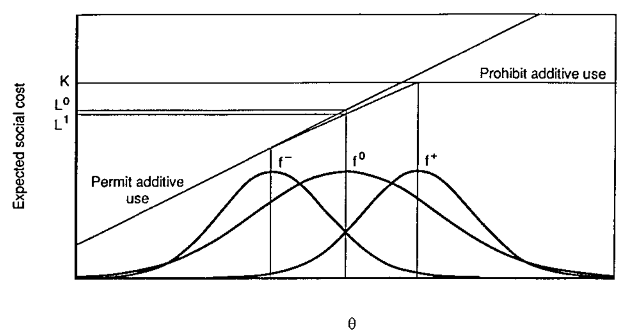

Fig. 5. Greater social cost of the substitute technology compared with Fig. 2 reduces the gain from prohibiting additive use conditional on learning it is more hazardous than currently indicated (f^+) and so increases L^1 and decreases EVI.

hazardous than prior information suggests is much smaller than in Fig. 2, and so the EVI is smaller as well. (If the risk from the substitute K exceeds the larger of the two posterior risk estimates for the additive L^+, the EVI is zero.) Similarly, perfect information about θ would be less valuable in this situation than in Fig. 2 because the probability that the new information would lead the regulator to prohibit additive use is smaller than in Fig. 2.

The fourth heuristic factor, relevance, applies when the preferred decision depends on the values of multiple uncertain parameters. Figure 6 illustrates the isopleths of a joint prior distribution for two uncertain parameters, θ_1 and θ_2, together with the boundary separating the regions where permitting and prohibiting the additive are the preferred decisions. Relevance determines the shape of the boundary between the two regions; in Fig. 6, the decision is more sensitive to the value of θ_1 than to the value of θ_2 and so improved information about θ_1 is more likely to alter the decision. When uncertain factors are multiplied together (as in the following example), all are equally relevant to the decision.

4.2. Example: Dichloromethane in Decaffeinated Coffee

The effects of the first three heuristic factors—uncertainty, informativeness, and promise—can be illustrated using an example from Hammitt and Cave (1991). Dichloromethane (DCM) or methylene chloride is a chlorinated solvent that was used to decaffeinate coffee until concerns about health risk led to substitution of alternative decaffeination processes. For simplicity, assume that the excess cancer risk to a representative consumer from drinking coffee decaffeinated with DCM can be represented as

$$\theta = \gamma\lambda\rho$$

where θ is the lifetime probability of developing cancer from this exposure source, γ is average daily decaffeinated-coffee consumption, λ is the DCM concentration in decaffeinated coffee, and ρ is the carcinogenic potency of DCM.[7]

Let the prior distributions for the three factors be lognormal. The excess risk θ is then lognormally distributed with parameters given by the sums of the corresponding factors for the three components as shown in Table II. (For θ and λ, which are bounded above by one, the lognormal distribution should be interpreted as an approximation to a logit-normal distribution.)

Assume the regulator is considering two options: to permit DCM use in coffee, leading to the uncertain risk described in Table II, or to prohibit use, in which case producers will substitute an alternative decaffeination process; assume the expected excess cancer risk if the alternative is used is $K = 8.3 \times 10^{-6}$ (and that uncertainty about this value cannot be reduced). If required to choose on the basis of minimizing expected risk and without additional information, the regulator should prohibit DCM use as the expected risk is 2.1×10^{-5}, slightly larger than the risk of the substitute (see Hammitt and Cave, 1991 for justification of the prior distributions).

The expected values of perfect information about the risk from drinking DCM-decaffeinated coffee and about each of the three factors are reported in Table II. The EVPI about the excess risk is 6.8×10^{-6}. The EVPI about the three factors contributing to uncertainty about the risk are monotonically related to the respective uncertainties: the EVPI for carcinogenic potency (6.4×10^{-6}) is greater than the EVPI for DCM concentration in coffee (1.6×10^{-6}) which is greater than the EVPI for coffee consumption (1.6×10^{-7}). The EVPI about potency is 93% of the EVPI about risk; because this factor dominates the uncertainty about risk, perfect information about

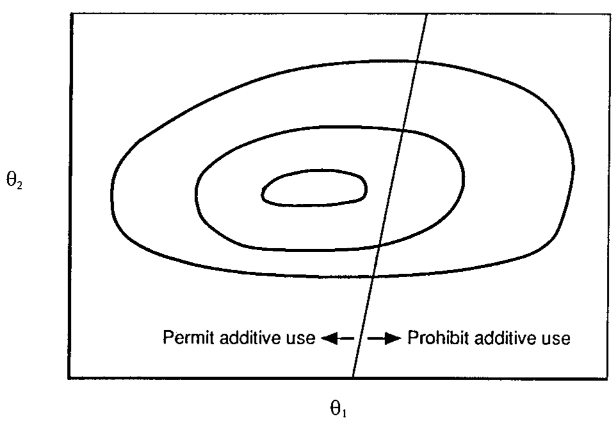

Fig. 6. Values of the uncertain parameters θ_1 and θ_2 for which permitting and prohibiting the additive are the preferred decisions and isopleths of the joint prior distribution for the parameters; improved information about θ_1 is more relevant than information about θ_2 as it is more likely to lead to a change in the decision.

[7] Note that this simple model could be extended to account for variability among consumers and coffees, nonlinearities in the relationship between consumption and internal dose, and other factors.

Table II. Prior Distributions and Expected Value of Information for DCM Risk

Component	$\log_{10}$ (median)	$\log_{10}$ (GSD[a])	Mean	Variance	EVPI[b]	SD(ω)[c]	EVI[d]
Consumption (γ)	5.5	0.25	3.7e + 5	7.1e + 10	1.6e − 7	0.1	1.0e − 7
Concentration (λ)	−7	0.5	1.9e − 7	5.2e − 14	1.6e − 6	0.5	6.4e − 7
Potency (ρ)	−5.33	1.25	2.9e − 4	1.3e − 4	6.4e − 6	10	6.5e − 9
Risk from DCM (θ)	−6.83	1.37	2.1e − 5	3.5e − 6	6.8e − 6		
Risk from substitute (K)			8.3e − 6				

[a] Geometric standard deviation.
[b] Expected value of perfect information about parameter.
[c] Standard deviation of measurement error.
[d] Expected value of information resulting from research about parameter with specified measurement error.

it is almost as valuable as perfect information about risk. In contrast, the EVPI about DCM concentration and coffee consumption are 24% and 2%, respectively, of the EVPI about risk. (In general, the sum of EVPI about individual components may equal, exceed, or be smaller than the EVPI about the components jointly.)

Obtaining perfect information about any of the factors is unrealistic. More plausibly, the regulator can conduct research that will yield improved information about the value of each of the factors. In this case, the EVI depends on both uncertainty about the factor and the informativeness of available research programs. In many cases, there is a tradeoff between the informativeness and cost of research; both are increased by sampling a larger number of units, for example. Following Hammitt and Cave (1991), we model research as an experiment that yields a measurement of the target factor multiplied by a random error term (analogous to measurement error); expressing this relationship in logarithms yields an additive error. For example, research on the concentration of DCM in decaffeinated coffee yields a value

$$x_\lambda = \log_{10}(\lambda) + \varepsilon_\lambda$$

where ε_λ is a random error term assumed to have moments $E(\varepsilon_\lambda) = 0$ and $\mathrm{var}(\varepsilon_\lambda) = \omega_\lambda^2$. Under these assumptions, the posterior distribution for $\log_{10}(\lambda)$ is normal with variance $1/(1/\sigma_\lambda^2 + 1/\omega_\lambda^2)$, where σ_λ^2 is the prior variance of $\log_{10}(\lambda)$.

The variance of the random error depends on the component under study. Because of the comparative difficulty in improving estimates of the three components the variance of the error is likely to be largest for carcinogenic potency and smallest for coffee consumption. Table II reports illustrative standard deviations of measurement errors for three research programs and the corresponding EVI, measured by the reduction in excess cancer risk for the representative consumer. For the selected parameter values, the EVI is greatest for research on DCM concentration and smallest for research on the carcinogenic potency of DCM. Even though prior uncertainty about potency is much larger than prior uncertainty about DCM concentration and coffee consumption, the greater accuracy with which concentration and consumption can be measured yield larger expected values of information about these components. Comparing the EVI for concentration and for consumption shows that, even though consumption can be measured more accurately, the EVI for concentration is larger because of the greater prior uncertainty.

In addition to uncertainty and informativeness, the expected value of information depends on the promise of the research program (as noted above, all three parameters are equally relevant because they are combined multiplicatively). Promise combines two attributes: the probability that a research program will yield information which is sufficiently persuasive to alter the regulatory decision and the gain (risk reduction) that results from altering the decision. In the example, the EVI depends on the risk K from consuming substitutes for DCM-decaffeinated coffee. As a function of K, the EVI about the risk from DCM-decaffeinated coffee and about each of the three factors is largest when K equals $E(\theta)$ and is smallest when K is either much larger or much smaller than $E(\theta)$. If the difference between K and $E(\theta)$ is large, it is unlikely that the regulatory decision that would be based on prior information is incorrect, and unlikely that research will alter information about $E(\theta)$ enough to alter the regulatory decision. If K is one or two orders of magnitude smaller than the value in the base case (Table II), for example, research results are unlikely to yield a posterior estimate of $E(\theta) < K$ and so the decision to prohibit

DCM use suggested by the prior estimate is unlikely to be altered. For $K = 8.3 \times 10^{-7}$ and $K = 8.3 \times 10^{-8}$, the EVPI about θ is 4.9×10^{-7} and 2.7×10^{-8}, respectively. These values are more than one and two orders of magnitude smaller than the EVPI in the base case.

5. COMBINING INFORMATION IN A PRIOR DISTRIBUTION

A prior distribution may be based on data for a similar parameter, the informed judgment of the risk analyst or subject-matter experts, or a combination of sources. Expert judgments may be encoded informally or through formal probability-elicitation methods. In the usual case, it is necessary to combine information from multiple sources, either to adjust the empirical distribution of a similar parameter to account for differences between that parameter and the one of interest, or to reflect differences among multiple experts whose judgments have been elicited using formal procedures. Many procedures for combining information have been proposed but none are clearly superior (Cooke, 1991; Jacobs, 1995; Clemen and Winkler, 1997). Moreover, the implications of alternative procedures when one is concerned to avoid overconfidence bias associated with the neglect of potential surprise has received little attention.

In the following subsections, we discuss two main classes of aggregation procedures, propose a method to adjust for overconfidence in expert judgments, and illustrate with an example using expert judgments about the magnitude of global climate change that would result from increasing atmospheric carbon dioxide.

5.1. Theory

Dependence among experts is both central to proper combination of expert judgments and difficult to evaluate. Judgments of multiple experts about a parameter can be extremely informative when those judgments are probabilistically independent, conditional on the "true" value. If, as is often the case, experts share much of the knowledge relevant to estimating a parameter value (e.g., a common scientific literature), the information contained in the union of multiple experts' judgments may be little more than that contained in a single expert's judgment (in effect, each expert may report an idiosyncratic perception of a consensus). Clemen and Winkler (1985) provide bounds on the number of independent experts whose combined information is equivalent to that of a larger number of dependent experts.

We consider two classes of algorithms for combining distributions across experts: weighted averages (opinion pools) and Bayesian combinations (Cooke, 1991; Jacobs, 1995; Clemen and Winkler, 1997). The weighted-average approach is simple, intuitively appealing, and can generate a wide range of combination rules. Bayesian approaches are motivated by treating each expert's judgment as data to be used in updating a prior distribution. A difficulty is that Bayesian methods require specifying the likelihood function or distribution of expert judgments conditional on the value of the uncertain parameter of interest. The limited available evidence on relative performance of alternative combination methods suggests that simple averages often perform nearly as well as the theoretically superior Bayesian methods (Clemen and Winkler, 1997).

A weighted average provides a transparent mechanism for representing unequal degrees of expertise. The weights can be assessed by the participating experts themselves or by others; they may also be used as sampling weights to represent the proportion of the universe of relevant experts with similar views. A weighted average of the experts' probability density functions (a linear opinion pool) can be motivated by the assumption that one of the experts' distributions is "correct," but it is not known which one; the weights represent the relative probability that each expert's distribution is "correct."

The simple weighted average can be generalized to provide a broad set of combination rules with potentially attractive properties, although it does not allow for convenient representation of dependence among experts' judgments. Following Cooke (1991), consider the case in which m experts provide subjective probabilities for each of n events; let p_{ij} denote expert i's probability of event j. A set of weights $\{w_i\}$ is assigned to account for differences in expertise or credibility among experts. Define the elementary r-norm-weighted mean

$$M_r(j) = \left(\sum_{i=1}^{m} w_i p_{ij}^r \right)^{1/r}$$

and the r-norm probability

$$P_r(j) = \frac{M_r(j)}{\sum_{k=1}^{n} M_r(k)}.$$

(These expressions may be generalized to the case of continuous random variables by replacing the summations with appropriate integrals.)

The r-norm probability provides substantial flexibility in combining experts' probability distributions. For $r = 1$, P_1 is the linear opinion pool. Taking the limit as $r \rightarrow 0$ yields

$$M_0(j) = \prod_{i=1}^{m} p_{ij}^{w_i}$$

so P_0 (the logarithmic opinion pool) is proportional to the geometric mean of the experts' distributions. For $r = -1$, P_1 is proportional to the weighted harmonic mean of the experts' distributions. In the limit as $r \rightarrow \infty$ and $r \rightarrow -\infty$, P_∞ is proportional to the largest of the probabilities offered by the experts and $P_{-\infty}$ to the smallest probability offered, respectively. For all values of r, the r-norm probability has the "zero-preservation property": if all experts assign probability zero to event j, then $P_r(j) = 0$. For $r = 0$, if any expert assigns probability zero to event j, then $P_0(j) = 0$ (Cooke, 1991).

An alternative approach to combining expert judgments is based on Bayes' rule. Beginning with a (possibly diffuse) prior distribution, the analyst treats each expert's distribution as new information and updates his prior using Bayes' rule. The details of the updating depend on the analyst's assessment of the likelihood function. The likelihood represents the probability that each expert will give the judgment he provides as a function of the underlying state of nature, and consequently incorporates information about the relative quality of experts' judgments (e.g., bias and overconfidence) as well as dependencies among the experts. The likelihood function provides a much more discriminating approach to characterizing the quality of an expert's judgment than does the single weight permitted by the weighted-average approaches.

Several models for the likelihood have been examined in the literature (Cooke, 1991; Jacobs, 1995; Clemen and Winkler, 1997). The geometric-mean combination P_0 can be derived from a model that treats each expert's assessment of a parameter as equal to the true parameter value multiplied by a random error (Mosleh and Apostolakis, 1986; Cooke, 1991).

Jouini and Clemen (1996) proposed a copula-based approach to combining distributions. A copula is a mathematical function that can be used to represent probabilistic dependence when coupling marginal probability distributions (the experts' judgments) into a multivariate distribution (the joint likelihood of the experts' judgments). This approach provides a flexible method for representing dependence among experts without restricting the form of the experts' distributions. In the following subsection, we describe the copula-based approach and use it to combine expert distributions.

Two properties that have been advocated for rules to combine expert distributions are the "marginalization" and "external Bayesianity" properties. Each property requires that the distribution obtained by combining expert judgments be invariant to specified procedural choices. However, most reasonable combination methods violate one if not both of these properties.

Marginalization requires that the combined probability of an event A be independent of whether the analyst combines the experts' assessed probabilities of A or sums the combination of the experts' assessed probabilities of $s > 1$ mutually exclusive and collectively exhaustive subsets of A. The r-norm probability satisfies marginalization only for $r = 1$.

Cooke (1991) argues that marginalization is a normatively important property, offering the following example: Two (equally credible) experts both assign probability 0.8 to the event that an old flashlight will not work. Using the geometric-mean combination P_0, the combined probability that the flashlight will not work is also 0.8. Assuming only two possible (mutually exclusive) failure modes, dead battery and corroded contacts, the probability of failure based on combining the assessed probabilities of each failure mode need not equal 0.8 if the experts disagree about the relative probabilities of the failure modes. If the experts assign probabilities (0.7, 0.1) and (0.1, 0.7), respectively, to the two failure modes, the combined probability that the flashlight will not work is 0.73 (the probabilities of dead battery and of corroded contacts are each $[0.1 \cdot 0.7]^{1/2,}$ that of no failure is $[0.2 \cdot 0.2]^{1/2}$). Because of the zero-preservation property, if the experts assign probabilities (0.8, 0) and (0, 0.8) to the failure modes, the combined aggregate probability that the flashlight will not work is zero, as the combined probability of each failure mode is zero (even though the experts agree that the marginal probability is 0.8).

In response to Cooke, one might argue that the experts' disagreement about failure modes provides important information about the credibility of each expert's probability that the flashlight will not work. Strong disagreement about the probability of the individual failure modes suggests that their agreement on the marginal probability of the flashlight not working is only coincidental and should have little significance for combining their judgments; Lindley (1985, 1988) argues that marginalization is not reasonably required of a combination rule. If the marginalization property is not required, however, the results of combining expert judgments may depend on the disaggregation of events and so the analyst will need to consider and justify whatever disaggregation is selected. Moreover, the marginalization property only applies to linear combinations of probabilities. When nonlinear models are used, combining judgments about the inputs will not necessarily yield the same result as combining model outputs even if the combination rule satisfies marginalization.

The external Bayesianity property applies to the case where one wishes to update a distribution when new data (or additional experts' judgments) are obtained. It requires that the same posterior distribution result whether one updates the prior distribution obtained by combining the experts' prior distributions, or updates each expert's distribution and then combines the resulting posterior distributions. Of the r-norm probabilities, only P_0 is externally Bayesian (Cooke, 1991). Lindley (1985, 1988) argues that this property is also not reasonably required of a procedure, but again if the combination rule is not externally Bayesian the analyst will need to justify the selected method for updating a combined distribution.

Acknowledging the possibilities of expert overconfidence and surprise suggests evaluating the effects of combination rules on tails and other subsets of the parameter domain assigned low probability. As already noted, the geometric-mean combination P_0 as well as other Bayesian combinations exhibit an extreme form of the zero-preservation property: if any expert assigns probability zero to some range of parameter values, the combined distribution does so as well. If experts' distributions are represented as triangular, uniform, or other forms that assign positive probability to only a finite interval, the interval to which no expert assigns zero probability may be small or nonexistent. This possibility represents a limitation of the Bayesian approaches but also highlights the danger in assigning zero probability to events that are not "impossible." We caution analysts to avoid assigning zero probability to parameter values that are not logically outside the bounds of the parameter (e.g., mass concentration ratios may be assigned zero probability outside the interval $[0-1]$). In addition, to reflect possible neglect of surprise events, it may be useful to extend the tails of experts' distributions as discussed in Sect. 2. (The results of extending the tails of experts' distributions before combining them and extending the tails of the combined distribution will not, in general, coincide.) The linear opinion pool may be preferred to a Bayesian combination because it represents greater uncertainty in the sense that its support (the set of parameter values assigned positive probability) is a superset of the support of a Bayesian combination. This follows because Bayesian combinations assign zero probability to parameter values for which any experts give zero probability, but the linear opinion pool assigns zero probability only to parameter values for which all experts give zero probability.

5.2. Example: Climate Sensitivity

Morgan and Keith (1995) elicited probability distributions for a number of parameters related to global climate change from 16 recognized experts in climate science. The elicitations were conducted using an extensive, formal process including pre-testing the elicitation instrument, advance distribution of background materials to experts, and day-long interviews. For illustration, we consider the experts' judgments for a central parameter in assessing the risk of climate change, the "climate sensitivity" ΔT_{2x}, defined as the equilibrium increase in global annual mean surface temperature that would result from a doubling of atmospheric carbon dioxide concentration from its pre-industrial level.

Morgan and Keith (1995) caution that it may not be appropriate to combine distributions across experts and that the diversity among the experts' distributions itself conveys important information about the degree of consensus in the expert community. In decision making, it may be better to analyze the problem using each expert's distribution individually to determine whether differences among experts' judgments are important to the decision and whether decisions that are robust to differences among experts' judgments can be identified (Morgan and Henrion, 1990). Nevertheless, a single decision must ultimately be made and so it appears useful to consider a distribution incorporating the best available information, which requires combining the judgments of relevant experts.

The experts' distributions for ΔT_{2x} are summarized by the boxplots in Fig. 7 (the participating experts are identified by Morgan and Keith (1995) but are not associated with particular distributions). Expert 4 gave two distributions, a base distribution and a second, wider distribution conditional on a "surprise" in understanding the climate system. With the exception of expert 5, who gave a very narrow distribution (minimum and maximum values of 0.06 and 1.0°C, respectively), most of the other experts' distributions are consistent with the values reported by the Intergovernmental Panel on Climate Change (IPCC): 1.5 to 4.5°C, with a "most likely value" of 2.5°C (IPCC has refrained from specifying the probability content of this range; Houghton *et al.,* 1990, 1996).

We combine the experts' distributions using the copula-based Bayesian combination and the linear opinion pool. Morgan and Keith (1995) provide no indication of the likelihood expert 4 assigned to the type of surprise for which his second distribution would be appropriate; for simplicity, we treat his two distributions equivalently (as if they were provided by separate experts).

Following Jouini and Clemen (1996), we use an Archimedian copula from Frank's family for combining the experts' distributions. An Archimedian copula treats the experts as exchangeable (i.e., the result is independent of which expert provided each distribution). Consequently, experts are treated as being

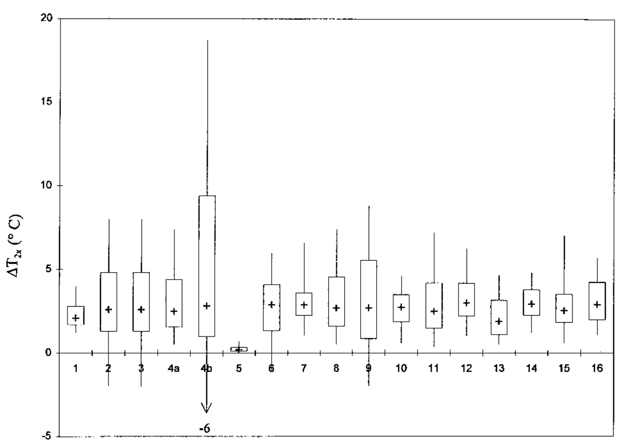

Fig. 7. Experts' probability distributions for climate sensitivity ΔT_{2x}, the equilibrium increase in global annual mean surface temperature that would result from an increase in atmospheric carbon dioxide concentrations to twice the pre-industrial level (Morgan and Keith, 1995). Boxplots illustrate median (+), interquartile range (box), 5th and 95th percentiles of the experts' distributions. Experts are identified by number. Expert 4 provided two distributions, "without surprise" (4a) and "with surprise" (4b).

equally reliable and the degree of dependence is the same across all pairs of experts.

Let $h_i(\theta)$ denote the subjective probability density provided by expert i (where $\theta = \Delta T_{2x}$) and let $H_i(\theta)$ denote the associated distribution function. The combined density function

$$f_n(\theta) = kC_{n|\alpha}[1 - H_1(\theta), 1 - H_2(\theta), \ldots, 1 - H_n(\theta)]h_1(\theta)h_2(\theta) \ldots h_n(\theta)$$

where k is a normalization constant and the copula is specified as

$$C_{n|\alpha}(u_1, u_2, \ldots, u_n) = \log_\alpha \left[1 + \frac{(\alpha^{u_1} - 1) \ldots (\alpha^{u_n} - 1)}{(\alpha - 1)^{n-1}} \right].$$

The parameter α characterizes the pairwise dependence among experts, with smaller values representing greater dependence. Independence is achieved in the limit as $\alpha \to 1$ and perfect dependence is achieved as $\alpha \to 0$.

To account for possible underestimation of extreme outcomes, we extend the tails of the experts' distributions using the compound distribution (Sect. 2) before combining distributions. The effect of extending the tails differs dramatically between the two combination rules. In addition, we examine the sensitivity of the combined distributions to including and omitting the atypical distribution provided by expert 5.

From each expert, Morgan and Keith (1995) obtained minimum and maximum values of ΔT_{2x} together with selected fractiles at probability increments of 0.1 or 0.05. The minimum and maximum values were not necessarily treated as absolute bounds, but the fractiles to which these values correspond are not reported (M. Granger Morgan, personal communication). For illustrative purposes we treat the reported minimums and maximums as the 0.0 and 1.0 fractiles.

We obtained probability density functions for the expert distributions by differentiating the cumulative distribution functions, which we approximated as linear between the elicited fractiles. Distributions with extended tails were obtained by replacing the 10% tails of the elicited distributions (i.e., outside the central 80% probability interval) with probability density proportional to $\exp(-|x|/u)$ where x is the difference from the expert's median normalized by the difference between the median and the adjacent quartile of the expert's distribution (the extended tails are asymmetric). Consistent with the analysis of errors in energy projections, we selected $u = 3$ (Shlyakhter *et al.,* 1994).

Combined distributions using the copula-based approach are sensitive to the treatment of expert 5's distribution, but are remarkably insensitive to the assumed degree of dependence among experts. Figure 8 illustrates using a value of α representing high dependence among the experts. For this value ($\alpha = e^{-18.2}$), the pairwise concordance probability is 0.9 (i.e., conditional on one expert overestimating ΔT_{2x}), the probability that each other expert overestimates ΔT_{2x} is 0.9, and so the probability that all 16 other expert judgments jointly overestimate ΔT_{2x} is $0.9^{16} \approx 0.2$). Combined distributions assuming smaller dependence among experts, including independence (pairwise concordance probability equal to 0.5), are almost identical.

Because the copula-based approach exhibits the zero-preservation property, the distribution obtained by combining all expert distributions assigns probability one to the interval (0.7, 1.0), as these are the only values of ΔT_{2x} to which all experts assigned positive probability. In contrast, the distribution obtained on combining all the distributions except expert 5's is consistent with the IPCC's stated range (1.5 to 4.5°C): the central 98% probability interval is (1.6, 2.8); the support is the interval (0.7, 4.9). Extending the tails of the experts' distributions sharply reduces the sensitivity of the combination to expert 5's distribution. The result is quite similar to that obtained by deleting expert 5's distribution from the combination except that it has a noticeably heavier left tail (on the interval [1.5, 2.2]) and assigns positive probability to the entire real line. In contrast, extending the tails has little effect on the combined distribution when expert 5's distribution is omitted; the two distributions cannot be distinguished in Figure 8.

Combined distributions based on the linear opinion pool P_1 (with equal weights) are presented in Figure 9 and compared with copula-based combinations assuming perfect dependence (pairwise correlation = 1.0) among the experts. The illustrated linear opinion pools are simple averages of the probability densities provided by each expert[8] after extending the tails to adjust for possible overconfidence. The results of combining the distributions without extending the tails are nearly identical. The linear opinion pool is not strongly sensitive to the inclusion or omission of expert 5's distribution. In either case, the result is very consistent with the IPCC range: the mode is near 2.5°C and most of the probability is assigned to the interval 1–5°C. Including expert 5's distribution adds a secondary mode near 0.5°C.

The copula-based distribution assuming perfect dependence among the experts and excluding expert 5 is remarkably similar to the linear opinion pool. Even under this extreme dependence assumption, the copula-based distribution remains very sensitive to expert 5's distribution; when it is included, the combined distribution shifts to the left and assigns very little probability to values of ΔT_{2x} greater than about 2°C.

Except when perfect dependence among experts is assumed, all of the (copula-based) Bayesian combinations are much narrower than the linear opinion

[8] Recall that the two distributions provided by expert 4 are treated as if they were provided by two of 17 total experts.

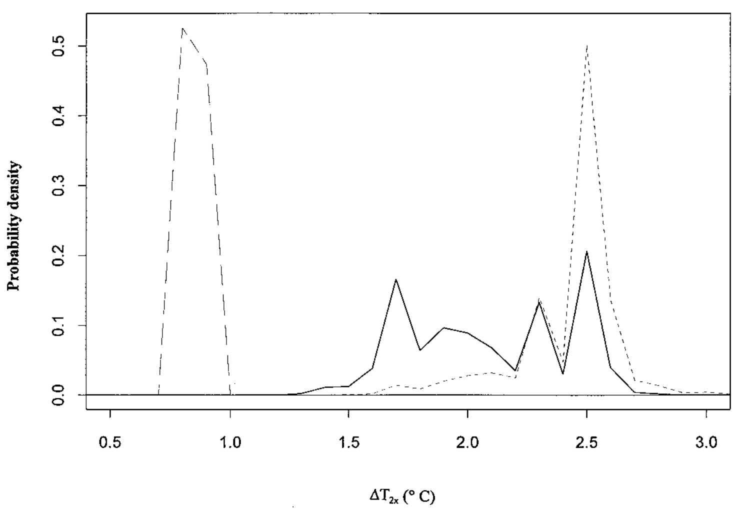

Fig. 8. Probability density functions for ΔT_{2x} obtained as copula-based Bayesian combinations of the expert distributions assuming high dependence among experts (pairwise concordance probability = 0.9, $\alpha = e^{-18.2}$). Long dashes: all experts; Solid line: all experts with added exponential tails; Short dashes: expert 5 excluded (densities with and without added exponential tails are indistinguishable).

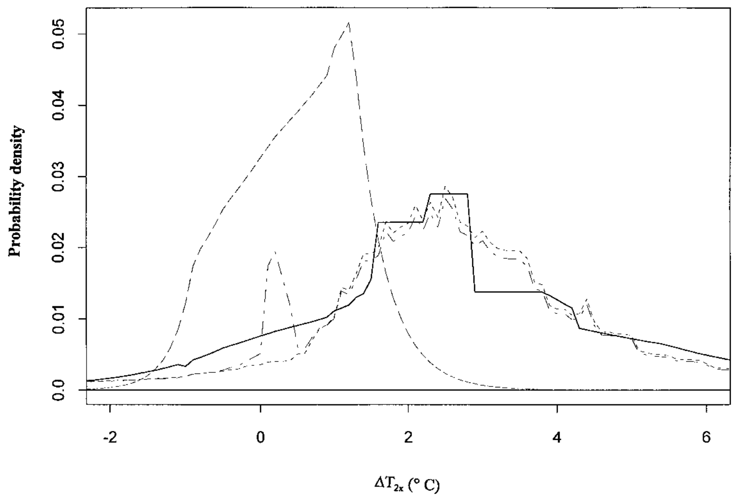

Fig. 9. Probability density functions for ΔT_{2x} obtained as linear opinion pools and as copula-based Bayesian combinations of the expert distributions assuming perfect dependence among experts. All combinations use expert distributions with added exponential tails. Mixed short and long dashes: linear opinion pool; Short dashes: linear opinion pool excluding expert 5; Long dashes: copula-based distribution with all experts; Solid line: copula-based distribution excluding expert 5.

pools and the IPCC uncertainty range (1.5–4.5°C). Probabilistic independence appears to be an untenable assumption in this case, since most of the experts have participated in developing the IPCC evaluations and all are familiar with (and contributors to) the literature summarized by IPCC. Even when a rather high degree of dependence is assumed, however, the combination of 16 or 17 expert judgments yields a much narrower range than most of the experts provided individually, reflecting the statistical power of multiple observations (even when the observations are not independent). The broad uncertainty interval produced by the linear opinion pool is only consistent with the Bayesian approach if an extraordinarily high degree of dependence is assumed.

Several alternative approaches to obtaining a distribution for climate sensitivity exist. For example, uncertainty about climate sensitivity can be examined by estimating ΔT_{2x} from multiple runs of general circulation and other models representing some of the processes through which CO_2 concentrations affect global climate. If one elicited judgmental multivariate distributions about the appropriate values of parameters to be used in such models and then derived the uncertainty in climate sensitivity by propagating these parameter uncertainties through a set of models, the result would differ from the results obtained by combining the distributions of climate sensitivity produced by the models.[9] These procedures would yield different results, even if the combination rule satisfied the marginalization property, because the climate models are nonlinear in most parameters. Alternatively, expert judgments could be characterized in terms of a climate-feedback parameter (Wigley and Schlesinger, 1985). The feedback parameter is proportional to the reciprocal of ΔT_{2x}; if the experts' distributions for ΔT_{2x} were transformed to distributions for the feedback parameter, combined, then transformed to yield a distribution for ΔT_{2x}, the results would also differ from the results of directly combining the expert distributions for ΔT_{2x}.

The appropriate level of detail at which to obtain expert judgment is unknown (and surely depends on context). There is some evidence supporting the view that experts should provide judgmental estimates about rather disaggregated components of a problem that can be aggregated using an appropriate model (Clemen and Winkler, 1997).

6. CONCLUSION

The assessed expected value of information depends on the prior distribution used to represent current information. Although exceptions can be readily generated, it is typically the case in risk assessment that the narrower the prior distribution, the smaller the assessed expected value of information. The well-documented tendency of individuals to be overconfident in summarizing their information, including particularly the tendency to underestimate the probability of surprise, can lead to large underestimates of the expected value of information. Such underestimates may be reduced by adopting prior distributions designed to mitigate the effect of neglecting potential surprise, such as a long-tailed compound distribution

[9] The expert judgments for ΔT_{2x} are based in part on the results of such model simulations.

calibrated to previous experience in the relevant domain.

In addition to uncertainty about one or more parameter values, the expected value of information depends on other characteristics of the problem, including the available decision options, research opportunities, and the resulting payoffs. Important factors may be heuristically represented as the informativeness of available research opportunities, the promise that research can be sufficiently persuasive to result in a decision that differs from the one that would be made in the absence of the research, and the relevance of uncertainty about a parameter to determination of which of the available decisions is superior.

In attempting to incorporate all the relevant evidence, analysts will often wish to combine information from multiple sources, including expert judgment. Numerous procedures for combining information across experts and other sources have been proposed but there is no consensus regarding the best procedure or even the relevant attributes for making that determination. The most appropriate procedure is surely dependent on the context of the problem and the state of the scientific theory, modeling, and data available for estimating important quantities.

The two prominent, simple procedures examined, Bayesian combination (using a copula to represent dependence among experts) and linear opinion pool, yield substantially different results and differ in their sensitivity to details of the experts' distributions. The linear opinion pool is less sensitive to individual distributions, and its results appear to be inconsistent with the theoretically justified Bayesian approach except under the extreme assumption of perfect dependence among experts. A potential weakness of the Bayesian approach is its extreme sensitivity to experts who assign zero probability to events that other experts view as possible; this limitation can be readily mitigated by extending the tails of expert distributions to reflect possible overconfidence.

Recognizing the possibility of surprise and adequately incorporating it in a risk or uncertainty analysis remains a challenge. The expected value of information is likely to be more sensitive to the probability of surprising outcomes than is the optimal decision under uncertainty, which suggests that research decisions in particular may benefit from thoughtful analysis of the characteristics of potential surprises. Analysis of possible surprises is likely to remain highly imperfect, however. By definition, potential surprises will remain elusive.

ACKNOWLEDGMENTS

Prepared for the "Second Workshop on When and How Can You Specify a Probability Distribution If You Don't Know Much?" Center for Risk Management of Engineering Systems, University of Virginia, Charlottesville, March 12–14, 1997. We thank the workshop participants (especially Robert Clemen), John Evans, John Graham, Charles Linville, Granger Morgan, Kimberly Thompson, and the referees for helpful comments, and Granger Morgan and David Keith for providing data. This work was supported by the U.S. Department of Energy (DE-FG02-95ER62109) and the Harvard Center for Risk Analysis.

REFERENCES

Alpert, M., and Raiffa, H., "A Progress Report on the Training of Probability Assessors," in D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, Cambridge, 1982), 294–305.

Chatfield, R. E., Moyer, C. R., and Sisneros, P. M., "The Accuracy of Long-Term Earnings Forecasts for Industrial Firms," *Quart. J. Business Econ.* **28,** 91–104 (1989).

Clemen, R. T., "Combining Forecasts: A Review and Annotated Bibliography," *Int. J. Forecast.* **5,** 559–583 (1989).

Clemen, R. T., and Winkler R. L., "Limits for the Precision and Value of Information from Dependent Sources," *Operat. Res.* **33,** 427–442 (1985).

Clemen, R. T., and Winkler, R. L., "Combining Probability Distributions from Experts in Risk Analysis," manuscript, Duke University (October 1997).

Cooke, R. M., *Experts in Uncertainty: Opinion and Subjective Probability in Science* (Oxford University Press, New York, 1991).

Dechow, P. M., and Sloan, R. G., " Returns to Contrarian Investment Strategies: Tests of Naive Expectations Hypotheses," *J. Financial Econ.* **43,** 3–27 (1997).

DeGroot, M. H., *Optimal Statistical Decisions* (McGraw-Hill Book Company, New York, 1970).

Gould, J. P., "Risk, Stochastic Preference, and the Value of Information," *J. Econ. Theory* **8,** 64–84 (1974).

Hammitt, J. K., "Can More Information Increase Uncertainty?" *Chance* **8**(3), 15–17, 36(1995).

Hammitt, J. K., and Cave, J. A. K., *Research Planning for Food Safety: A Value-of-Information Approach* (R-3946-ASPE/NCTR, RAND, Santa Monica, CA, 1991).

Hawkins, N. C., and Evans, J. S., "Subjective Estimation of Toluene Exposures: A Calibration Study of Industrial Hygenists," *Appl. Ind. Hyg. J.* **4,** 61–68 (1989).

Henrion, M., and Fischhoff, B., "Assessing Uncertainty in Physical Constants," *Am. J. Phys.* **54,** 791–798 (1986).

Hilton, R. W., "The Determinants of Information Value: Synthesizing Some General Results," *Mgmt. Sci.* **27,** 57–64 (1981).

Houghton, J. T., Jenkins, G. J., and Ephraums, J. J., (eds.), *Climate Change: The IPCC Scientific Assessment* (Cambridge University Press, Cambridge, 1990).

Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A. and Maskell, K. (eds.), *Climate Change 1995:*

The Science of Climate Change (Cambridge University Press, Cambridge, 1996).

Huang, C. F., and Litzenberger, R. H., *Foundations for Financial Economics* (North-Holland, New York, 1988).

Jacobs, R. A., ''Methods for Combining Experts' Probability Assessments,'' *Neural Comput.* **7,** 867–888 (1995).

Jouini, M.N., and Clemen, R.T., ''Copula Models for Aggregating Expert Opinions,'' *Operat. Res.* **44,** 444–457 (1996).

Kahneman, D., Slovic, P., and Tversky, A., (eds.), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, Cambridge, 1982).

Lichtenstein, S., Fischhoff, B., and Phillips, L. D., ''Calibration of Probabilities: The State of the Art to 1980,'' in D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, Cambridge, 1982), pp. 306–334.

Lindley, D. V., ''Reconciliation of Discrete Probability Distributions,'' in J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (eds.), *Bayesian Statistics* (North-Holland, Amsterdam, 1985).

Lindley, D. V., The Use of Probability Statements,'' in C. A. Clarotti and D. V. ''The (eds.), *Accelerated Life Testing and Experts' Opinions in Reliability* (North-Holland, Amsterdam, 1988).

Morgan, M. G., and Henrion, M., *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (Cambridge University Press, Cambridge, 1990).

Morgan, M. G., and Keith, D. W., ''Subjective Judgments by Climate Experts,'' *Environ. Sci. Technol.* **29,** 468A–476A (1995).

Mosleh, A., and Apostolakis, G., ''The Assessment of Probability Distributions from Expert Opinions with an Application to Seismic Fragility Curves,'' *Risk Anal.* **6,** 447–461, (1986).

Murphy, A. H., and Winkler, R. L., ''Reliability of Subjective Probability Forecasts of Precipitation and Temperature,'' *Appl. Stat.* **26,** 41–47 (1977).

National Risk Council, *Understanding Risk: Informing Decision in a Democratic Society* (National Academy Press, Washington, D. C., 1996).

Presidential/Congressional Commission on Risk Assessment and Risk Management, *Framework for Environmental Health Risk Management: Final Report* (Washington, D.C., 1997).

Raiffa, H., *Decision Analysis* (Addison-Wesley, Reading, MA, 1968).

Raiffa, H., and Schlaifer, R., *Applied Statistical Decision Theory* (Harvard University Graduate School of Business Administration, Boston, 1961).

Rothschild, M., and Stiglitz, J. E., ''Increasing Risk: I. A Definition,'' *J. Econ. Theory* **2,** 225–243 (1970).

Shlyakhter, A. I., Kammen, D. M., Broido, C. L., and Wilson, R., ''Quantifying the Credibility of Energy Projections from Trends in Past Data: The U.S. Energy Sector,'' *Energy Policy* **22,** 119–130 (1994).

Shlyakhter, A. I., ''Improved Framework for Uncertainty Analysis: Accounting for Unsuspected Errors,'' *Risk Anal.* **14,** 441–447 (1994).

U.S. Department of Energy, Energy Information Administration, *Annual Energy Outlook with Projections to 2010,* DOE/EIA-0383(92), (Washington, D.C., 1992).

Wigley, T. M. L., and Schlesinger, M. E., ''Analytical Solution for the Effect of Increasing CO_2 on Global Mean Temperature,'' *Nature* **315,** 649–652 (1985).

Winkler, R. L., and Poses, R. M., ''Evaluating and Combining Physicians' Probabilities of Survival in an Intensive Care Unit,'' *Mgmt. Sci. 39,* 1526–1543 (1993).