Risk Analysis, Vol. 30, No. 3, 2010 DOI: 10.1111/j.1539-6924.2010.01361.x

Commentary

Ignorance Is Not Probability

William A. Huber*

The distinction between ignorance about a parameter and knowing only a probability distribution for that parameter is of fundamental importance in risk assessment. Brief dialogs between a hypothetical decisionmaker and a risk assessor illustrate this point, showing that the distinction has real consequences. These dialogs are followed by a short exposition that places risk analysis in a decision-theoretic framework, describes the important elements of that framework, and uses these to shed light on Terje Aven's criticism of nonprobabilistic purely "objective" methods. Suggestions are offered concerning a more effective approach to evaluating those methods.

KEY WORDS: Decision theory; interval analysis; subjective probability

1. DIALOGS ON RISK ASSESSMENT

A Decisionmaker (DM) has asked a Risk Analyst (RA) for advice. She has two prospects to choose from, both risky. Each offers an opportunity to gain \$6. She believes prospect A has close to a 1/6 chance of success. Specifically, this means both the DM and RA view this prospect as acting like a lottery in which six tickets are entered, each with an equal chance of winning, and one of them belongs to the DM. According to all the information available to them, prospect B has somewhere between a 0 chance and 2/3 chance of gain: these bounds are based on theoretical limits, not on any data or experience. Let us listen in on the conversation that ensues.

1.1. The Initial Dialog

DM: Which prospect is better, A or B? I must choose one of them.

RA: Well, the statistical expectation of prospect A equals 1/6 times \$6, or \$1. Prospect B has a statistical expectation, too, and I know it is somewhere between \$0 and \$4, but I don't know which.

DM: How does that help me? It seems you are telling me that B is both better and worse than A.

RA (channeling Daniel Ellsberg):[1] You're right. Let's look into this. We can model prospect B as a lottery, too. Let's suppose there are many tickets entered and some of them are yours. In fact, we know that up to two-thirds of them have your name on it. As a concrete example with numbers that are easy to calculate with, suppose there are 90 tickets.

DM: According to my calculator, that means as many as 60 of those tickets are mine.

RA: Right. But, in the spirit of giving you impartial advice, I should let you know that maybe none of them are yours.

DM: None!? Then I definitely don't want to choose B.

RA: Slow down. We don't know that none are yours. It's just as possible that 30, or even 60, have your name on them. That would give you much better chances by choosing B. In fact, if only 15 of the tickets—one-sixth of the total—were yours, then both prospects would look the same, because you would have a 1/6 chance of winning in both.

DM: I see. So if fewer than 15 of the tickets in B are mine, I should stick with A, but if more than 15 are mine, I should prefer B.

RA: Right.

*Address correspondence to William A. Huber, Quantitative Decisions, 1235 Wendover Road, Rosemont, PA 19010; tel: (610)-527-3599; whuber@quantdec.com.

DM: I can count for myself: I could have zero, one, or up to 14 tickets and that's 15 possibilities that would make me favor A. Or I could have 16, 17, and up to 60 tickets and that's $60 - 16 = 44$ possibilities that would make me favor B. And in the remaining case of 15 tickets it's a toss-up.

RA: Close. There are actually 45 possibilities that favor B. Remember, there are 61, not 60, possibilities in all.

DM: OK, I'll believe you. Higher mathematics was never my strength. But since the 45 possibilities that favor B outnumber the 15 ones that favor A, my decision is clear. But I hired you also to provide justification: give me some support here.

RA: You're using the Principle of Insufficient Reason. It says you can compute the chances by weighing the possibilities for and against an action. It has a long history and has been used by many famous people.

DM: That helps a lot: my boss likes polysyllabic incantations repeated over time immemorial by accepted authorities. Oh, one other thing: you said prospect A has a statistical expectation of $1. What is the expectation of B?

RA: I can set up a probability tree for that. Let's see, the first branch is the number of your tickets and the second branch is the outcome of the lottery. So there's a 1/61 chance you have no tickets: you definitely lose in that case. There's a 1/61 chance you have one ticket, giving you a chance of one in 90 of winning. And . . . well, there's a lot of math to do here because we have to look at 61 different cases. Let me go back to the computer and I'll get you an answer soon.

DM: Can the computer do that?

RA: Sure, I'll set up a Monte Carlo calculation. That's a way to simulate situations like this where the exact computations are mathematically intractable.

DM: That sounds sophisticated. Go ahead and let me know the answer.

RA (later, calling from the office): I ran 10,000 trials to simulate prospect B. You won the lottery in 3399 of them. That means you should win about 34 percent of the time, giving an expectation of $2.04.

DM: It sounds like B is twice as good as A. Thank you! By the way, we might have multiple independent opportunities at prospect B, maybe three of them. I'm worried about the worst case. What are the chances we could lose in all three?

RA: I ran that in the simulation too and it came out to be around 63%.

1.2. The Follow-Up

Has the Risk Assessor acquitted himself well? Is prospect B truly twice as good as A? Let's look at one scenario. Time passes. RA has not heard from DM since she paid his bill. Pleased with the advice he gave her earlier, but recognizing the limits of his understanding, he has been studying Bayesian statistics in the interim and longs for a chance to apply these methods. The telephone rings and DM is at the other end. After the preliminary niceties are exchanged, they get back to business:

DM: You recall I chose prospect B.

RA: Yes—I calculated it was twice as good as A.

DM: Well, we lost.

RA: I'm sorry to hear that. But you knew it was a risk, right? Losing once doesn't show that you made a bad decision. It was just the luck of the draw.

DM: You're right. That's why I needed to justify my decision beforehand. I'm glad you had the Principle of Insufficient Reason up your sleeve. Did you know we have another crack at prospect B? Do you think we should go for it again, or should we choose A?

RA: Let me work it out. We used an uninformative prior last time, which is a Beta(1, 1) distribution, and that . . .

DM (interrupting): Hold on! Did you just say you gave me *uninformed* advice?

RA: No, no, not at all. That was just a way of saying we assumed all the possibilities between 0 and 2/3 were equally likely. Now we actually have some real data: you tried B and lost. That updates our prior to a Beta(1, $1 + 1$) distribution, whose mean is 1/3. That's your personal probability distribution for the fraction of those 60 tickets bearing your name. It tells us you should act as if you have 20 tickets in the game. Your chances of winning are 20/90 for an expected value of $1.33. That's less than $2.04 but still greater than $1. You should choose B again.

DM: If you say so. I'm a little suspicious of all that "insufficient" and "uninformative" mumbo-jumbo, but numbers don't lie. I'll go with B again. Thank you.

As it transpired, DM lost again and consulted once more with RA, who lowered his estimate of B's expectation to $1 (again using a Bayesian calculation). In fact, the chances in A were slightly lower than 1/6, so prospect B still narrowly won out over A, and DM plunged a third time, losing yet again. How could this happen? It could be bad luck, which RA had estimated as a 30% possibility at the outset. Alternatively, it could be that *prospect B was truly and consistently bad*. It could behave as if none, or very few of the 90 tickets, were in DM's favor.

In the scenario given, it appears that DM's interests were not well served by the probabilistic analysis offered by RA: DM acted under the assumption that she had better than even chances—63%, actually—of succeeding at least once, whereas in fact her chances of success were practically nil. Somehow, the analysis failed her. In other possible scenarios the chances of prospect B might have been considerably greater than 1/3. *Not being aware of that could have created lost opportunities for DM.* Thus, replacing pure ignorance—not knowing what the true chances of the prospect were—creates additional risks: on the one hand, risks of making bad choices, and on the other hand, risks (perhaps unknowable or unquantifiable) of not being aware of just how good some prospects are. This all came about because DM pushed RA to express this ignorance with a probability distribution instead of accepting RA's first response, which was to provide a range of winning chances.

1.3. An Alternative Scenario

Now consider how the risk assessment could have played out if RA had not given in to the pressure to offer a prior probability distribution for DM's chances in prospect B. Let us return almost to the beginning of the initial dialog.

> *RA:* Prospect B has a statistical expectation, too, and I know it is somewhere between $0 and $4, but I don't know which.
>
> *DM:* How does that help me? It seems you are telling me that B is both better and worse than A.
>
> *RA:* I am, but I can't help that: we simply don't know where we stand with B. I see how that can frustrate you as a decision maker, so let's talk about what you can do. One way to think about it is to contemplate the various possibilities. How would you feel if B actually had very low chances of success but you went with it anyway? How would you feel if B had high chances of success but you instead chose A (and watched as some daring competitor ventured B, perhaps)?
>
> *DM:* Obviously I can afford to lose, for otherwise I wouldn't even be considering prospects with such low odds. But you're not helping me choose here!
>
> *RA:* The potential difference in expectation between these prospects is as great as $4 minus $1, which is pretty big. That would be worth investigating. Do you have resources and time to get a little more information about B?
>
> *DM:* No, I don't. However, I may have several opportunities at B or prospects exactly like B. Does that help?
>
> *RA:* Yes, it does. If you're willing to invest a few times in B in order to learn more about it, we can use your experience to decide whether B or A is a better choice.
>
> *DM:* How many is "a few"?
>
> *RA:* In the worst case, three. For if by then you haven't gained anything from B, we will figure its actual chances aren't anywhere near 2/3 or even 1/3 and are likelier equal to 1/6 or less, which will make A more attractive.[1] Otherwise, you will already have received a return on your investment, your decision will look good, and you can keep choosing B even longer, always with the option of going to prospect A if B doesn't pan out in the long run.
>
> *DM:* I'm sorry, I can't afford to do that. But I appreciate knowing where we stand. I'll go back to my directors and let them know we simply don't have a basis for choosing between A and B and if they want some basis, they will have to pay to get more information. It's a pity we couldn't employ you to do those high-powered calculations everybody is talking about, you know like in Las Vegas?
>
> *RA:* You mean Monte Carlo, I think. But that's ok: my job is to give you an objective impartial assessment of the information you have and to help you translate that into the best possible decision for you.

2. COMMENTS

2.1. Decision Theory

Former *Risk Analysis* editor Robert Cumming, in this journal's inaugural editorial,[(2)] characterized the purpose of risk assessment as "to help in the decision-making process." He went on to write: "Risk analyses *will* be made. If decisions, based in part upon these analyses, are to be optimized, it is important that they be as impartial and as accurate as it is possible to make them." This places risk assessment squarely within a well-established decision theory framework: decisions have uncertain consequences, with their attendant losses and gains, and these consequences are not perfectly predictable. They are modeled as outcomes of a random variable described by a probability distribution that usually is not known, but about which some information is available in the form of data and background knowledge. This probability distribution is often called a "state of nature." Optimal decisions are those for which the statistical expectation of loss (the familiar probability-weighted consequence) is best. This gives rise to many closely related concepts, of which three merit attention here: how the probability distribution is described (its "parameters"), the data, and data-driven guesses about what the parameters might be (their "estimates"). Prospect B of the dialogs

[1] These are Bayesian conclusions. A classical analysis would yield somewhat different advice and be couched in more complex language.

exemplifies these: its outcome is modeled as a multiple of a Bernoulli random variable, the parameter chosen to describe it is the expectation, and the data are DM's sequence of losses.

2.2. Parameters

So-called frequentist (or "classical") and Bayesian (or "subjective") statistical methods differ over how the parameters are modeled. In both paradigms knowledge of the parameters is mediated through data, which are related in various given ways to outcomes of the random variable. The potential values of the parameters may be constrained by assumptions and theory. The Bayesian paradigm exploits additional knowledge (or assumptions) about the parameters by supposing that they, themselves, are random variables. Their probability distributions are called *prior* distributions. The key argument against using Bayesian methods is that often there is no mechanism or no apparent justification for linking prior distributions to empirical facts.[3] When a prior distribution is wrong and few data are available, decisions are likely to be bad.

Prior distributions are frequently determined by considering the uncertainty in the mind of the decisionmaker. Different decisionmakers with differing degrees of uncertainty could use different priors and conceivably then arrive at different decisions. This is a basis for calling priors "subjective." However: "All statistical methods that use probability are subjective in the sense of relying on mathematical idealizations of the world."[4] Thus, subjectivity is an epithet validly applied to all quantitative models. Nevertheless, let us not lose sight of Cumming's exhortation to make risk analyses *as accurate as possible*: when a prior carries accurate information about the true state of affairs (as encoded in the distribution parameters), its use can lead to more accurate risk assessments and better decisions. Otherwise, its use can degrade the accuracy and lead to bad risk assessments.

Both statistical paradigms are capable of representing epistemic uncertainty. Classically, knowledge is captured by the data, the structure of the model, and the model's assumptions: what kinds of probability distributions might govern the outcomes and what are the possible ranges of their parameters. Lack of knowledge is implicit in the fact that more than one probability distribution is possible. (If not, we are no longer dealing with a statistical problem, but merely assessing odds in a gamble.) This framework is richer and more flexible than it might seem, considering that the probability model is usually multivariate (which captures statistical interactions in the data) and can include within its workings any quantitative relationships suggested by theory. Bayesian methods make the lack of knowledge even more precise by assigning a specific prior distribution to the set of possible parameters. If that additional precision leads to better decisions, that is good; otherwise, we should not use a prior. The choice of methodology is not (or at least should not be) a philosophical one: it is entirely pragmatic.

2.3. Data

Nonprobabilistic approaches to modeling uncertainty often begin with the data. To see how, consider a recent monograph by Ferson *et al.*[5] (whose 2007 publication date reveals it as a better source of information than the 1996 paper, by many of the same authors, to which Terje Aven is responding[6]). It is explicitly motivated by "incertitude," which "arises in several natural empirical settings" in which the data are "intervals rather than point" values, such as binned data, censored data, and data that have been digitally rounded. Ferson *et al*. show how to adapt standard statistical computations to interval-valued data by means of interval-valued arithmetic. This is not so much a replacement for probability theory as it is a piggy-backing on top of probability theory. It is done in a spirit of "if we know only that the data could lie in certain specified intervals, then let's consider what decisions we would reach for all possible values of those data that are consistent with those interval constraints." If you want to use classical methods for decision making (Aven's "objective" risk assessment) then you get an interval-valued version of those methods: a range of decisions consistent with all that is known explicitly about the data. If you use Bayesian methods (Aven's "subjective" risk assessment), then you get a range of Bayesian decisions consistent with all that is known about the data *and which incorporate all the "scientific judgment about the unknown quantities" you care to include.*

2.4. Estimates

The important thing to keep in mind about parameters is that they are purely theoretical constructs: they are neither measured nor observed. We make guesses, more politely known as estimates, as to what their values might be. A statistical estimator is just some procedure by which data (modeled as outcomes of the random variable in question) are converted into statements about the parameters. Typical statements are of the form "the parameter has such-and-such a particular value," "the parameter lies within such-and-such a range of values," or—only in the Bayesian case—it can be a probability distribution for the parameters (the posterior distribution).

2.5. Analysis

Terje Aven's criticism of interval analysis and other nonprobabilistic methods in risk assessment rests on a dichotomy of risk assessment purposes:

(1) "Objectively" describe "unknown quantities."
(2) "Obtain a scientific judgment" about quantities "from a qualified group of people.... [A] picture of what is known or not known about a particular issue is created."

Although Aven does not tell us exactly what he means by "quantities" and "issue," his examples suggest that these are states of nature, such as the "unreliability" of a system of "units" in a process plant.

Taking Cumming seriously, we should be asking ourselves how to tell when the "picture" in (2) is accurate. What assurance do we have that a "scientific judgment" of probabilities—which are notoriously difficult for people to estimate well—*that is made in addition to and independently of the data* bears any trustworthy relationship to empirical reality? Indeed, requiring that the people have "strong competence in the field" as well as training in probability estimation is specified in the hope they will be objectively correct in their judgments. Evidently, we care about expert opinions to the extent those opinions have *empirical* validity. This establishes a connection between the objective purpose (1) and the subjective purpose (2). However, relying on scientific judgment, expert elicitation, and prior probabilities stands out as a solution that is decidedly nonscientific in spirit. Although this form of subjective information-gathering is essential in many applications, it should be considered a method of last resort, not as a preferred mode of conducting risk assessment. Focusing on validity and accuracy as desirable properties of a risk assessment should motivate us to look for direct information about the facts and to exploit scientific judgment in other ways first. Excellent alternative uses of expert scientific opinion are in the construction and criticism of the model itself, determining what data to collect, and how best to conduct their measurement.

To this one might reply, "that's all well and good: risk assessment should use science when it can, and that indicates looking to nature rather than human authority for answers whenever possible. But what else can you do when there is lack of knowledge or no data at all?" What exactly can "lack of knowledge" or "no knowledge" mean? In this context, it implies *the experts have no basis on which to offer a prior probability distribution anyway*! But Aven, if I am reading correctly, tells us "they should assign numbers also in this case" because the decisionmaker insists. Such a capitulation is an error: Wovon man nicht sprechen kann, darüber muß man schweigen.[2] Assigning numbers in the absence of evidence does not serve the decisionmaker, but rather deceives her. As the dialogs illustrate, the deception is serious: it can lead to manifestly poor performance in decision making. Interval analysis is motivated by a desire to avoid this mistake while accommodating DM's need to exploit all information possible and striving for Cumming's ideal of accuracy.

2.6. Resolution

A better way to conduct a critical evaluation of interval analysis, possibility theory, or any of the other nonprobabilistic methods proposed for risk assessment is to assess how well they are likely to perform. Because interval analysis has been proposed for use within a standard decision-theoretic framework, it is amenable to standard methods of evaluation. Specifically, when an unbiased statistical estimator is applied to interval-valued data using interval arithmetic, does it remain unbiased? When confidence level procedures are applied, do the resulting intervals (or intervals of intervals) maintain their requisite levels of coverage? When hypothesis testing is applied to interval-valued data, do the resulting tests have the intended size and power? The methods to answer these and related questions are well established and well known:[3] we only need adapt and apply them to get the answers.

There nevertheless is common ground here. When we allow that the techniques of interval analysis apply equally well to classical and Bayesian methods, Aven's dichotomy vanishes and accommodating "subjective" scientific judgment (in the form of prior

[2] "Whereof one cannot speak, thereof one must be silent."[7]

probability distributions) becomes possible within a risk assessment that uses interval analysis. If there is any remaining conflict, it lies in the interpretation of interval-valued *data*, not in any matter of philosophy or risk assessment methodology. (Specifically, one can question whether treating binned data, censored data, *etc*. as true intervals fully captures the information inherent in such data. Often it does not—but this is not the place to pursue such issues.) Also, both Aven (at least in places) and the works to which he is responding seem to agree that where you have no basis to provide probabilities, do your best to get additional information—accurate information—about them. Regardless, we need to know the limitations of our data and models and we need to present them fully and correctly.

REFERENCES

1. Ellsberg D. Risk, ambiguity, and the savage axioms. Quarterly Journal of Economics, 1961; 75(4):643–669.
2. Cumming RB. Is risk assessment a science? Risk Analysis, 1981; 1(1):1–3.
3. Kiefer JC. Introduction to Statistical Inference. New York: Springer-Verlag, 1987.
4. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis, 2nd ed. Boca Raton: Chapman & Hall/CRC, 2004.
5. Ferson S, Kreinovich V, Hajagos J, Oberkampf W, Ginzburg L. Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty. Setauket, NY: Applied Biomathematics, 2007. Available at http://www.ramas.com/intstats.pdf. Accessed on 1/5/2009.
6. Aven T. On the need for restricting the probabilistic analysis in risk assessments to variability. Risk Analysis, 2010; 30(3):354–360.
7. Wittgenstein L. Tractatus Logico-Philosophicus. Routledge: Abingdon, Oxon., 1922. (Original with English translation by C.K. Ogden.).