“The Pebble Remains in the Master’s Hand: Two Careers Spent Learning (Still) from John Evans”

Adam M. Finkel and George M. Gray

This paper has been prepared for the Harvard Center for Risk Analysis “Risk Assessment, Economic Evaluation, and Decisions” workshop, September 26-27, 2019
**Introduction:**

“Of course,” the title refers to the forgettable 1970s TV series “Kung Fu,” where the student was not allowed to leave the monastery until he could snatch the pebble from the master’s hand—as soon as he did, he had to pack up and move on. The two of us left the “monastery” anyway, sans pebble, but are still learning from John.

John started many of us on a lifetime of learning about risk assessment and management. In an increasingly irrational world, his mantra was always “analysis is useful” (along with “mice are more like rats than people are... in most cases”).

But John has always stood for, and advanced, a *brand* of analysis that is not merely useful because practitioners say it is, or because it has more (deserved) appeal than “qualitative risk assessment” or the kind of analysis needed in support of “precaution.” When all one needs to implement a policy is “noun plus adjective,” as in “[name of chemical here] BAD” or “Expenditures BAD,” the only “analysis” necessary is to assert that exposures to the substance, or the completely analogous “exposures” to the costs of control, could be non-zero under some scenario. John’s career has stood for the premise that these attributes of analysis make it useful:

- Careful attention to uncertainty and to interindividual variability, keeping the two very different phenomena conceptually and mathematically separate, but combining them when enlightening (in particular, the extent to which any citizen can know what risk she faces is limited both by the uncertainty in anyone’s risk and by the partial or complete inability of analysis to tell her where she falls on the distribution of interindividual risk);

- Full quantitation of exposures, risks, valued benefits, and control costs. We can’t pin any of these down precisely, but that’s no excuse for reducing quantitative

---

1 Environmental Risk Management: Is Analysis Useful? Air Pollution Control Assoc., Pittsburgh, PA, 1986. John Evans was the first author of this volume.
information to yes/no pronouncements, or to “green/yellow/red” bins—instead, we should quantify the *uncertainty* whenever we quantify the quantities.\(^2\)

- “Analysis is useful” because it exists in service of better *decisions*. Analysis that exists merely to extrude more and more information-in-a-vacuum, divorced from any consideration of how the information can/should affect *action*, is vain, in both senses of that word.

- John has made us all think hard about the virtues of design/specification rules versus numerical targets.\(^3\) We *think* he agrees with us that while we can “move the dial” on exposure until the marginal benefit of further reductions equals the marginal cost of more controls (that is, a performance standard dictated by cost-benefit balancing), this alone doesn’t get us anywhere unless we understand how “the dial actually gets moved.” The technologies are lumpy/discrete, and so what we really should be doing is using risk and economic information to compare *real choices* that are available to us. But John has also helped us remember that it’s often too facile merely to advocate for “BAT” or “ALARA”—if we have the capability to reduce risks, at ever-increasing costs, far below *de minimus* levels, we should think hard before insisting that society do so.

**Discussion:**

This paper discusses four vexing problems in risk-based decision-making that John has shed massive light on over the last nearly 40 years, and has perennially challenged the two of us and others to think about. Our theme is that work remains to be done on each of these, but that some of that work would merely involve listening to ideas John has already offered.

---

\(^{2}\) John and I (A.F.) “naturally” assumed in our 1987 paper on the value of information (JAPCA) that risk was uncertain but that cost wasn’t. This was naïve of us, and I’ve written several papers since then arguing that cost uncertainty is larger, but far more well-hidden, than risk uncertainty.


DRAFT
1) **Thresholds are irrelevant to decision-making unless they occur at relevant exposures.**

A thriving industry continues to attack the assumption that “low” doses of a substance will pose some risk if “high” doses are clearly risky. The belief that all toxicants (carcinogens and/or non-carcinogens) must have a threshold challenges not only the time-honored assumption that risks at low doses are proportional to those at high doses (low-dose linearity), but also the less health-protective assumption that risks rise more steeply at high doses than at lower ones. In other words, if a dose-response relationship has a threshold, it may be irrelevant that effects are seen at “high” doses, and therefore any positive epidemiology or toxicology study should be discounted, ignored, or deemed “interesting” but not an indication of human risk. If adopted as science-policy, this stance would, of course, completely upend much of the fields of toxicology and epidemiology as they relate to chemical, radiological, and perhaps biological exposures.

The concept of the threshold has considerable merit, both for very low exposures to carcinogens (if faithful DNA repair exceeds the rate of new DNA lesions) or non-carcinogens (if, for example, mucociliary clearance can completely remove infrequent trespass by fine particles). But from the under-appreciated but absolutely fundamental point of view of decision theory and risk management, the existence of a “threshold somewhere” is completely unimportant to any decision effecting reductions in exposure from a point clearly above the threshold to one “above but less far above” said threshold. 4

In other words, a regulator should be unimpressed with ANY claim of threshold behavior unless it could possibly affect the magnitude of risk at the specific “low” doses to which she wishes to regulate. For this reason, the frequent claims seen in the threshold literature that “one molecule cannot possibly be harmful at all” make no sense—when government seeks to reduce lifetime exposure to a substance from (say) $10^{26}$ molecules to $10^{25}$, the behavior of the dose-response function at 1, 10, or ten

---

4 We emphasize that there are two situations in which a threshold might not be crossed: the one mentioned here (where pre-decision and post-decision exposures are both above the threshold), and the obverse case, in which pre-decision exposures are already below the threshold). This latter case could occur where the exposures seen in bioassays are never experienced by humans, or where the exposures seen in epidemiologic studies are no longer encountered anywhere.
billion molecules is not worth a moment’s thought (or a decade’s delay in taking action while advocates investigate this behavior). As a theoretical matter, we are happy to assume that the generic dose-response function (in the absence of specific information to the contrary) is not “linear no-threshold,” but “linear PLUS threshold”—an approximately straight-line response in the frank-effect range and somewhat closer to the origin, and a threshold somewhere between zero exposure and “low” exposure. But we suggest that the burden of providing evidence of a decision-relevant threshold falls at the feet of those who dispute the utility (not the veracity, but the utility) of a linear model from “high” to “less high” doses.

In upcoming work, we plan to first systematically and completely evaluate all peer-reviewed articles claiming that any particular toxicant has a threshold, by partitioning the set of articles into those which do—and which do not—attempt to quantify where the threshold occurs. It may be the case that the vast majority of all claims of a threshold make no attempt to argue that the threshold matters for any conceivable policy decision; if this is true, then decision-makers and the public need to understand that the health benefits of modest exposure reductions being proposed would likely be the same whether or not the dose-response had a “threshold somewhere.” The same conclusion would hold if most of the subset of (the remaining) studies that do claim to pinpoint a specific threshold have failed to acknowledge that this threshold is (far) below any level to which society currently contemplates regulating.

We will then proceed to examine individually the (smaller) subset of studies that do claim there is a threshold at a decision-relevant level, and see to what extent these claims rely merely on “common sense,” theoretical arguments, and/or on the inability to observe effects at a level then pronounced to have zero risk. To the extent that a claim of a policy-relevant threshold merely results from lack of power in the underlying toxicology or epidemiology, we will estimate the lower statistical bound on each substance’s “threshold” based on the power of the data to rule out a linear or sub-linear dose-response, and evaluate whether this more conservative “threshold” in fact remains relevant for any completed or pending decision about the substance.
2) There is too great a focus on human health reference values from “authoritative bodies” that work too slowly and sometimes on the wrong things.

When potency (and hence risk, and hence risk-based control) values will be assumed to be non-existent (zero) until sufficient evidence accrues for an authoritative body to carry out an assessment, all incentives flow in the direction of making it harder and harder to agree on those values. This complicates many decisions – especially those that involve chemical substitutions and the like, and simply sets up a risk treadmill as we move from one problem to the next. This is the “missing risks” problem—and we urge that EPA’s Integrated Risk Information System (IRIS) and the other “potency exercises” switch gears from the “gold or nothing” standard to a “provisional first; gold second” process where “10 year risk assessments” are done to improve provisional potency values, rather than being a precondition for having any official potency estimate.

Our current process for developing official human health reference values (HHRVs), such as those from IRIS, can take years or even decades to complete and potentially endanger public health in the meantime. The slow pace of review leaves many potentially dangerous chemicals without risk values needed for good public health decisions. Even when they are done, they are challenged by stakeholders, NAS committees and many others. We need something John Evans has advocated for decades, namely, faster ways to use existing information to generate risk values, even for chemicals with little toxicologic data. And when this is done, we need to reflect the uncertainty in these values to help with decisions and guide future research5.

Not having authoritative HHRVs often means the potency of chemicals is implicitly treated as zero. This means important risks may be missed while the very slow gears of official assessment grind and potentially sets up a move from one problem to another. When manufacturers wanted to remove bisphenol A (a chemical with authoritative values and

hence in the spotlight) from their products they turned to unassessed chemicals with similar properties. An EPA evaluation demonstrated that for use in thermal printing paper the substitutions may have increased, rather than decreased, risk, based on assessing the risk of the substitutes using data currently in hand.

Approaches to developing provisional (or rough and ready) HHRVs based on in vitro tests, structure activity modeling and empirical relationships exist now. Many of these have been around for a long time and are designed to help provide numbers useful for regulatory decisions. Other approaches seek to use short-term data to predict points of departure for chronic risk assessment. These are usually independent of the specific toxic effect which we know does not predict well across species anyway. It is very true that these estimates will be uncertain, and a real challenge is developing methods to characterize that uncertainty. As John would say, uncertainty does not mean ignorance and we can use the information available to help avoid the “missing risk” problem of unassessed compounds and provide better information for chemical management decisions.

Perhaps the greatest challenge is getting people comfortable with using these rough and ready HHRVs. Decision makers will have to contend with uncertainty in an explicit way. Other stakeholders will likely object too. For many, a lack of authoritative risk values is a feature, not a “bug,” and the current slow and contentious approach avoids scrutiny. Using alternative methods to develop HHRVs means the default position will be that all chemicals

---


pose some risk which may need to be managed. Toxicologists will cry that decisions are being made without full testing of chemicals and advocates will focus on possible sensitive populations or specific endpoints that might be of concern.

Despite these concerns, implicit treatment of no authoritative HHRV as zero risk makes for bad decision analysis. We urge the development, perhaps by the current authoritative bodies but perhaps by new groups, of provisional potency values, with their attendant uncertainty, to ensure that public health decisions are well informed. Tools like VOI analysis (below) can then be used to characterize whether, and what, new data might need to be gathered to take revise a provisional HHRV.

3) **Two fundamental desiderata in risk analysis and management may be incompatible:** the desire to fully characterize uncertainty, versus the desire to apply reasonable assumptions and models and avoid paralysis.

There is “plenty” of uncertainty (and variability) to characterize, and to explore the decision ramifications of, in the typical appraisal of a contaminant in the environment. The precise shape of the dose-response function is uncertain, the slope at any region of the function depends on sample-size uncertainty in the observed effects in a bioassay or epidemiology study (and in the latter, on uncertainty in the exposures of each study group), the conversion of exposures from rodents to humans depends on a host of uncertain pharmacokinetic parameters (or on an allometric equation of $[\text{body weight}^x]$ whose exponent is highly uncertain), the concentration of the contaminant at any location-time coordinate is uncertain, the ingestion/inhalation/absorption by residents is uncertain, and so on.

But the *model uncertainty* (MU) in the foregoing is assumed away: what about the substantial additional uncertainty contributed by the possibility that the proposed exposure reduction crosses a biological threshold? What about the possibility that effects seen in test animals are irrelevant to humans, or that elevated incidence rates in human

---

studies are confounded and not caused by the exposure? What about the possibility that the health effects associated with exposure are treatable and therefore not grave? *Does it violate a basic principle of uncertainty assessment to analyze risk and uncertainty* conditional on a set of assumptions about causality, evidence, and relevance, or must we acknowledge (all) possible alternative assumptions and widen our uncertainty bounds because we can’t be sure our assumptions are correct?

There are two completely reasonable ways to account for fundamental model uncertainty: (1) articulate and array all of the plausible models and derive risk estimates (with their parameter uncertainty and variability!) for each; or (2) fully articulate a system of which default models will be used (and why!) in the absence of compelling evidence to the contrary. The first is more faithful to the desire to avoid overconfidence; the second may be more efficient and realistic in light of incentives and strategic behavior.

Both ways pose difficulties. In order for the first way to truly differ from the second, it must provide weights (subjective estimates of the degree-of-belief to be assigned to each model)—otherwise we would be left with multiple and incompatible risk estimates with no way to either combine them into a single uncertain estimate (you need weights for that) or to assess the consequences of acting as if the wrong model was correct (you need weights for that too, otherwise all one can say is “we might be very wrong, with unknown likelihood”). So without subjective weights, the first approach becomes the second approach—acting as if our model is correct, without an estimate of how likely that model error is. But the assignment of weights to incompatible models, though it has been accomplished and refined over many iterations, is controversial and frequently criticized for being arbitrary and easily-manipulable.

Needless to say, the second way (choosing a set of default assumptions and models, and replacing one or more of them with an alternative model/assumption upon some kind of showing of greater plausibility in a specific case) also has profound difficulties associated with it. How can we profess to care about uncertainty if we deliberately censor some of it? Who decides when an alternative model is sufficiently superior to a default that it becomes
the one viable way to estimate risk for that situation/substance? What happens to the “minority” possibility that the original default predicted for risk when it is supplanted by an alternative?

As a highly simplified example of the contrast between the two approaches, suppose the risk from Substance X would have one mean-and-variance estimate if the most common rodent tumors found in its bioassay were truly relevant to humans, and a much lower distribution of possible risk if analysts agreed that these tumors were irrelevant and a different tumor type was the most precautionary dataset from which to estimate potency. The impulse to include both assumptions comes from the desire to be honest that either is a possible representation of the truth; the impulse to use one assumption until the other is deemed superior comes from the desire to avoid paralysis, to avoid “monkeywrenching” (if the goal of some stakeholder is to lower the likelihood of regulation or lower the likely regulation’s stringency, then an arena in which “dueling models” vie for portions of the subjective weight between 0 and 1 is one in which it’s more profitable to complain about the weights—or to sponsor misleading research—than to actually resolve the scientific controversy).

On the one hand, two successive committees of the National Academy of Sciences (cite Blue Book, Silver Book) strongly recommended that EPA and the other EHS regulatory agencies continue to rely on a system of defaults/departures, but that they should make much more clear which defaults they initially rely on (and why!), what level of “proof” (e.g., a scintilla more likely than the default, “clearly superior” to the default, beyond a reasonable doubt correct) they will require to discard a default in favor of an alternative, and how a proponent of an alternative model/assumption can show (quantity and quality of evidence, bearing on specific research questions) that the alternative is in fact slightly/significantly/unequivocally superior to the default. But on the other hand, EPA in particular (over the last 10 years or so) has moved towards a position of accepting neither defaults/departures nor model averaging—instead, it claims a goal of “reducing reliance on defaults,” but instead purports to choose the “best” models and assumptions de novo with each risk assessment. This merely replaces largely precautionary defaults—assumptions DRAFT
that sometimes have support from decades of research and dozens of experiments in clearly analogous situations—with a different assumption chosen, in effect, “by default” because the Agency is eager to accept preliminary research on supposed “mode of action” and unwilling to regard all the prior information as relevant. A long footnote in the Silver Book criticized this evolution at EPA, arguing that the goal should be to “reduce reliance on incorrect or inapt defaults, not to repudiate all defaults for no valid reason, apparently to no avail.

We do not necessarily fully agree about which of the two ways is better, only to point out that MU is of great importance, but that it is possible to pay too much attention to it (or the wrong kind of attention) as well as too little. We do, however, fully agree that if multiple models are to be considered for a given step in the risk assessment equation, there is a right way and a very wrong way to do so. To declare that (this is a simplified case of the general case) “the risk is either very large (let’s call that risk X) with probability $p$ or very small (let’s call that risk zero) with probability $(1-p)$,” because one of two incompatible theories is right and the other is wrong, is a reasonable but dangerous place to start. One thing the quoted utterance does not mean is that “the risk is $p$ times X.” Model averaging should not become “risk averaging,” but can be very useful as “decision averaging.”

In other words, the right way to handle “the risk is either huge or tiny” is to use that information to compare the performance of two or more decision options. We need to consider what we might gain or lose IF the risk is huge but we make the decision that makes sense for a tiny risk, versus the gain or loss if the risk is tiny but we act as if we know it is huge. This can’t be done anywhere near correctly by acting as if the risk is the weighted sum of $[p \text{ times huge}]$ and $[(1-p) \text{ times tiny}]$. What is the harm from a $[1-p]$ probability of “over-reacting,” versus the harm from a $p$ chance of “under-reacting”? is the right question, and it can’t be answered without considering the costs and benefits of each decision under each state of nature. John taught us both that back in the days when the computing power

---

of an iPhone would have filled a stadium full of mainframes, and yet it is still not followed today as the truism it is.

And we offer one other set of significant cautionary notes about “a full treatment of MU.” The main attraction of incorporating MU rather than relying on defaults is that the former approach may allow decision-makers and the affected publics to “see the full light of uncertainty.” If incorporating MU carries the baggage of subjective weighting, delay, and possible “decisions guaranteed to be wrong,” and doesn’t even fully depict uncertainty, then it may be a marginal improvement with substantial downsides. We suggest here that there is much more to MU than simply supplementing, overturning, or “watering down” precautionary and reasonable defaults with other reasonable interpretations of MOA, interspecies scaling, and the like.

What would a full treatment of MU in risk, benefit, and cost look like? We will attempt in future work to provide a typology, with examples, of model uncertainties that few if any risk and cost-benefit analyses ever consider. Certainly there are alternative exposure and fate-and-transport models that are rarely considered alongside the traditional ones. Ditto with the way we currently handle uncertainty in the VSL (as the central tendency of many stated-preference experiments or revealed-preference studies, but rarely incorporating the model uncertainty that makes it difficult to choose one type of study over the other). But we believe that is the economic aspects of cost-benefit analysis that are severely handicapped by the tacit and pervasive use of unacknowledged default assumptions without appreciation of MU. For example, regulatory economics routinely assumes, without explicit mention, that partial-equilibrium cost estimates are good surrogates for over general-equilibrium ones, that technological learning and/or economies of scale are unimportant, that price rises will reduce demand rather than spur demand for substitute goods, etc.

And even within the realm of dose-response modeling, there are ways in which we censor important uncertainties. For example, suppose an epidemiology study shows a RR with confidence limits going from [0.8 – 4.0]. That is a classic “negative” result because the LCL
is below 1. But why shouldn’t we include the (let’s say) 50% chance that the RR is > 1 in our risk/uncertainty estimation? Our objection to including “fringe” assumptions and giving them expert-derived weight is really no different from using the entire confidence interval on “negative” bioassay and epidemiology results.

Finally, most of the concerns raised here about risk estimation apply equally to hazard identification. Systematic review is supposed to free analysts from the constraints of defaults and allow for a full treatment of uncertainty for causal inference. But we argue that since it can’t possibly do that, we should compare it to the prior approach (hazard identification based on defaults) with eyes open to how much the more complicated and time-consuming approach can truly deliver in terms of fully explicating MU.

4) Decision-makers who refuse to require value-of-information calculations to guide the choice between action and analysis (and to guide the contours of any further analysis) ought to admit they are walking around with sunglasses on when the skies are dark.

John Evans has been a strong advocate of the use of Value of Information (VOI) analysis for many years. Sometimes in its formal sense, as the expected gain that would come from reducing the uncertainty in estimates of the consequences of alternative choices in a decision. Often, though, John would advocate merely stopping to think more qualitatively about how valuable new information would really be in making choices. He would encourage the enormously useful thought-question “how much would the information have to change my risk estimates to make me change my mind?” Often, it is difficult to imagine any experiment or survey or sampling effort that would make a big enough difference to change a choice.

John is not the only one who has encouraged the use of VOI information in environmental, health and safety decisions. For example, Committees of the National Research Council of
the National Academies of Science have urged EPA\(^\text{13}\) (Science and Decisions) and FDA\(^\text{14}\) to expand the use of VOI in making research and information gathering investments. There are a wide range of technical papers and reports that identify ways in which VOI could be applied by agencies\(^\text{15}\).

However, VOI, in any formal sense, has not caught on at all in the regulatory world. There are both implementation issues and technical challenges that seem to have stymied its use. Perhaps the biggest implementation issue is that in order for VOI to be applied risk estimates need quantitative estimates of uncertainty, something regulatory agencies have very rarely developed. In addition, any formal decision analytic tool needs to specify \textit{a priori} the options being considered. It is clear that many in decision-making position are uncomfortable stating the options under consideration in advance. There also seems to be a general belief that uncertainty analysis and VOI are too difficult for decision makers to understand and use appropriately – a view we believe reflects badly on the decision makers, not the analysts\(^\text{16}\).

There are technical challenges to using VOI in EHS risk decisions too. One of the greatest is actually knowing how much information a given data collection event will deliver and how much it will reduce uncertainty. Much early VOI focuses on the expected value of perfect


\(^\text{15}\) e.g, Dakins, M. (1999) \textit{The Value of the Value of Information}. Human and Ecological Risk Assessment: An International Journal, 5:2, 281-289  

information, but we know that an animal toxicology experiment, exposure assessment, or cost of implementation survey will provide only partial information for an analysis. Estimating how much uncertainty will be reduced with different sources of information is a continuing challenge. It may also be difficult to know how much it will cost to deliver a specific piece of information. Some information may be generalizable across decisions and would therefore be even more valuable than assumed for a single choice. These and other technical issues, while they need to be considered, are not obstacles to the use of VOI today.

Perhaps a place we can start is with John’s question of bounding the magnitude of uncertainty reduction necessary to change a decision. For example, imaging an abandoned hazardous waste site with chemicals identified as carcinogens in the soil but no contamination of groundwater. There are three decision options: Do nothing, put a cap of clean soil above the contaminated dirt, or dig up and remove the contaminated soil. A risk assessment, using default procedures like linear dose-response relationship between exposure and cancer risk and standard dust exposure assumptions, finds the lifetime cancer risk to be $1 \times 10^{-3}$. At this risk level the relatively best remediation choice is to remove the soil. If the risk were $1 \times 10^{-5}$ the choice would be capping, and if under $1 \times 10^{-6}$ the choice would be do nothing. Because the remediation options are “lumpy,” John’s question would have us ask if and how new information could even move us from one to another. In this case, if the assumption of linear dose response is a major source of uncertainty, we would have to think that we could do an experiment that would leave no more than a 1% chance that the true dose-response is linear, in order to change our choice from soil removal to capping. It is highly unlikely that any information gathering would yield this level of precision, so this form of VOI thinking has helped solidify a decision. In other cases, it may be that readily-obtainable information, on cost of alternatives or exposure profiles, could indeed matter and more formal analysis would be called for.

Those who make research and data gathering decisions, intended to help guide and improve decisions, but don’t use VOI approaches are likely to squander resources and miss opportunities to shape better choices.
Conclusions:

These four themes may seem disparate, but they are inter-related and all hearken back to the answer John provided in the APCA volume in 1986: “analysis IS useful.” Earlier we emphasized that analysis is more useful, perhaps only useful, when it is done quantitatively, with careful attention to uncertainty and variability, and in ways that allow feasible choices to be compared along multiple dimensions. But more importantly, we suggest his 1986 title was pointing the field towards a probing examination of what “useful” means. Of course good analysis has value and utility—but we don’t merely want the analysis because it is a tool; we (should) want the results of our actions to be “useful.”

So perhaps the real question behind the 1986 question was and is “are unanalyzed actions useful?”—and John and we his students would say “no.” For an easy target, consider the first 24 launches of the Space Shuttle between 1981 and 1986. Arguably (insert citations here), NASA risk managers did not heed the results of the risk assessments done there, and while the missions before the Challenger disaster were useful, they were not optimal. On the other hand, John set both of us on careers that included substantial time in regulatory agencies, where we can look back on decisions we made affirmative but also on equally weighty decisions we made by failing to decide, waiting to decide until after we could no longer influence policy, or changing the subject in order to “make” a decision about some other problem (thereby making a decision about the problem we were shunting or punting). So we have learned, from working with John and from life after we “left without the pebble,” that whether one sees one’s self as an analyst or a decision-maker (a somewhat arbitrary and unfortunate bifurcation), one needs to be relentless in asking as often as possible “what did I decide today?” If the answer is “I decided we weren’t ready to decide,” John’s work challenges us to then ask, with great humility but also with great urgency, “what are we waiting for, and why?”

So all four of the topics we discussed here tell the same story, with variations. Analyzing dose-response data (toxicologic, epidemiologic, or both) to categorize a stressor as “threshold or not” can be valuable, but only if knowing which is which is expected to
improve an outcome. Using provisional HHRVs can seem deflating, *but only if* refining them is expected to improve an outcome. “Reducing reliance on defaults” can increase real or perceived sophistication, *but only if* doing so is expected to improve an outcome. And if any of these refinements matter, which they certainly often will, some form of VOI analysis is waiting in the wings to provide the answers to questions like these three.