[01/31/2003; @@@ QM Master inc fig]

HOW "DECISION AID" CAN MISLEAD DECIDERS DUE TO CONFLICTS OF INTEREST

Rex V. Brown
School of Public Policy
George Mason University

ABSTRACT

Quantitative aids often squander their potential to improve decisions. The logic of an aid may be sound enough, but if it fails other essential tests of usefulness (such as sound input and appropriate output) a trusting decider can be worse off than before. One major reason for the failures is that the aider's priorities may conflict with the decider's. The aid may employ models of little use, because they are tractable and interesting. It may disregard important knowledge, because it cannot be validated. It may make unrealistic assumptions, because they are convenient. The decider may have to take control of the aiding process to protect his interests; by specifying what the aid is to do, by checking the assumptions it is based on, and by directing how it is to be used. The argument is illustrated with examples drawn from extensive consulting experience.

1 INTRODUCTION

1.1 Problem.

Decision makers in government and elsewhere make judgments all the time about what is happening in the world and what to do about it. However, they are not always confident that they are making soundest use of what they know or could learn—and with good reason. The need for better decisions of all kinds is widely accepted, particularly now that stakes rise with globalization, and technological complexity challenges informal judgment. For example, individuals make mistakes that blight their careers (Baron, 1998); companies fail and governments fall because of "organizational foolishness" (March and Shapira, 1982; Allison, 1971).

So a decider, D, turns to formal aids that are intended to enhance his judgment, in the sense that it better serves his objectives (when, for example, a manager has to commit company resources). The aid's contribution may be in the form of indicating a preferred choice (e.g., tools of decision analysis and operations research). Or it may be in the form of reducing D's uncertainty about a key factor in the decision, like option impact (tools of statistical sampling and probability theory).

There are other kinds of aid not addressed here. "Decision aids" may be employed for purposes other than helping someone to make up his own mind. In particular, they may be intended to advocate or defend a position to someone else[1]. I am only concerned here with making D's choice more *rational* in serving *his* interests. A slanted analysis may be very useful for *advocating* his rational choice to others, but not itself be rational.[2] Interestingly, the consulting demand for such "persuasion aiding" may exceed real aiding.

1.2 This paper

Scope. Technically advanced aids (both qualitative and quantitative) have been available for more than 50 years, but with limited practical success (see below). This paper explores the hypothesis that practical progress in aiding has been hindered, not by technical immaturity (which time should cure), but by more durable behavioral phenomena that seriously impair application. A number of these phenomena have been addressed elsewhere (Thomas 1989, Brown 1970).

However, one of the most serious impediments appears to have been largely neglected. It is the motivation of the technical *decision aider, A*. He may report to D directly or through a support service, or work for an outside contractor, or act as an independent consultant, but many of his priorities appear to be common.

---

[1] A Navy agency asked me to evaluate whether Congress should appropriate money for bombers or aircraft carriers. I accepted, provided the Air Force be informed of any findings. The request was withdrawn.
[2] E.g. by considering only the criteria that favor the preferred option.

Many As are not appropriately motivated (from D's point of view) for reasons that may persist without outside intervention. I will explore what influences A to disregard those interests, with what effect, and what D can do about it. Aids, in this paper, are evaluated in terms of the extent they serve the interests of whoever actually makes the decision. Those interests may differ from those of, say, shareholders or society at large, important though these be (Brown 2000).

The focus here is on quantitative decision aiding (abbreviated to QDA or "aiding"). It is often in the form of decision analysis, arguably the most practically oriented aid and the one in which I have most experience. That experience, covering 40 years of varied policy consulting and aiding, is the basis of my conclusions, which may be no more than promising hypotheses.

1.3 *Hierarchy of aid "clients"*

Effect of an intermediary "client". The complete aiding service that D uses often includes an intermediary agent (such as an in-house support group or an independent research contractor) for whom A works. This "client" will have his own motivations distinct from those of the D (say turf preservation or follow-on business), which may influence A and impair the aid he produces.

Higher-order aid "clients". D may himself report, nominally or implicitly, to a higher body, whose interests he is supposed to serve—but may not[3]. For example, A company CEO may set up offshore partnerships to make him rich at the cost of shareholders he "reports" to (as in the notorious Enron case). I will not address whether, ethically, the A *should* serve an D with suspect interests.[4]

1.4 Structure of paper.

The next section discusses the past record of practical aiding, and relates it to a varied list of essential requirements that an aid must meet to be useful. Section 3 presents and illustrates the central hypothesis on linkages between aiding strategy and aid usefulness, involving A's motivations and their impact on meeting aid requirements. It is expressed in the form of a causal net.

The next two sections flesh out this hypothesis based on practical experience, and document it with case illustrations of pathways through the net. Section 4 focuses on individual aid deficiencies and Section 5 on individual A motivations. Section 6 draws

[3] A nuclear regulator told me that, when deciding what reactor safety measures to require, it would be embarrassing for him to acknowledge publicly in a formal model that he took into account political "hullabaloo" (of the type caused by the Three Mile Island accident), as well as legitimate health effects (apparently negligible at TMI). His A did not take this into account and so understated risk (as D valued it)

[4] A student in a junior high decision analysis class shared his dilemma on whether to become a drug dealer with his teacher (really!). If it appeared that the student's own best interests, as he evaluated them, supported his drug dealing but society's interests did not, what should the teacher as a responsible decision "A" do (Brown, in press)? I hold that in exceptional cases a civic-minded teacher can jettison professional standards for the greater public good and mislead the student into not dealing. Ron Howard finds that "unprincipled".

some conclusions, including about action implications for Ds and higher-level policy makers and confirmatory research still needed.

2 USEFULNESS OF QUANTITATIVE DECISION AID

My experience as an advisor to executives on how to make effective use of *other* people's aids has given me the opportunity to observe and evaluate many case studies from the perspective of the customer. It has been a chastening experience (Porter, 1989).

2.1 The QDA record

"Above all, do no harm" (Hyppocrates)

Half a century ago, QDA, especially operations research, was justly credited with having done much to win World War II. It was then widely believed that QDA could go on to better the welfare of mankind in general. Public and private organizations started to use QDA extensively. Businesses in particular trusted the analyses enough to let them overrule their own judgment and drive their action. However, the results were sometimes catastrophic (as in the "car parts depot" example below).

QDA tools themselves have certainly improved greatly since then, both in scope and technical quality, but so far they appear to have had *limited* practical success. What *has* changed over the years is that users have become more cautious about acting on recommendations based on QDA.

Consider the following cases.

- A $4m probabilistic analysis showed that a US reactor was one of the safest; but an NRC regulator put it on a "watch list" of dangerous plants.
- A car company, acting on an operations research study, closed down parts depots; but had to reopen them when the remaining depots proved insufficient to handle demand.
- A probabilistic UN study found a negligible chance that any country could illegally build nuclear weapons without detection; but Israelis disagreed and bombed an Iraqi reactor.
- A DOE report based on decision analysis selected a short list of three potential nuclear waste sites; but the Secretary of Energy chose a different three.
- A statistical survey reported that 61% of households lowered their thermostats to save energy; but policymakers assessed only 30% when choosing conservation incentives.

We may be tempted to interpret such disconnects between "aided" and "unaided" judgment as encouraging evidence that Ds need aids to avoid blunders (Polister 1991, von Winterfeldt and Edwards 1986). However, I believe that the contrary is commonly true and that *unaided* judgments can actually be sounder, including the above cases. This view is controversial (Howard 1992), but the following anecdote is telling. An American political scientist (whom I prefer not to name) evaluated the quality of

decisions by Israeli generals in the Six Day War by comparing them (unfavorably) with his own analysis of the same decisions (based on data available at the time of the war). I strongly suggest that no responsible Israeli citizen would put any military decision in the hands of a foreign academic rather than of his own generals[5]. Unfortunately, misguided use of aids is not always so readily dismissed.

2.1.1 Pessimistic views of QDA

Other authors have cautioned against over-reliance on QDA (Quade 1980; Majone, 1980; House, 1988; Moore, 1973; Strauch 1980). A blue ribbon NAS Committee on Risk Analysis and Decision Making found that only a tiny fraction of QDA potential was being realized (Simon, 1988). Three surveys of the usefulness of decision analysis in business and government, over a thirty-year span, found few cases where D's claimed a beneficial impact (Brown 1970; Ulvila and Brown 1982; Brown 1987a).

Moreover, there are widespread reports that QDA consulting and QDA teaching at professional schools are not expanding significantly[6]. Watson, who co-authored a major decision analysis text (Watson and Buede 1987) later cast doubts on its usefulness (Watson 1996). Jackson Grayson, a pioneer in applying decision theory (Grayson 1960), cautioned against uncritical use of management science in general after he himself became an influential D[7] (Grayson 1978).

A leading decision scientist is said to have been asked if he were using QDA to resolve on a personal dilemma. "Certainly not!" he replies, "*This is an important problem!"* This tale, whether true or not, is widely retold in the aiding community, which suggests that its message resonates even there—though not comfortably.

2.2 The potential of QDA

I do not share this gloom. Properly used, I believe QDA can greatly benefit society. The problem is that the *potential* for QDA to make decisions sounder has barely been tapped. Aids certainly *can* do good, and often do. The Institute of Operations Research and Management Science regularly publicizes success stories[8], documented by testimonials from satisfied clients. QDA offers Ds valuable logical discipline to sloppy thinking that can improve their judgment[9]. By "improved" I mean judgment that takes fuller account of available knowledge, and thereby provides a more realistic, and therefore useful[10],

---

[5] Particularly if the academic were basing his judgment on an inevitably incomplete reconstruction of what the generals knew, was the case here.

[6] Even at Harvard Business School, the pioneer in teaching decision analysis to MBAs, it is no longer required.

[7] As head of the Federal Price Control Board

[8] Reported annually in the Interfaces in connection with the Franz Edelman award for successful OR applications.

[9] I distinguish the roles of three types of "user": D, client and beneficiary, who may have different agendas (if they are not the same party). The "D" ultimately makes the decision (e.g. government executive); the "client" (e.g. research agency) employs the A, on behalf of the "beneficiary" (e.g. the general public).

[10] By "realistic" I mean a judgment that promises to be veridical, i.e. to correspond to how the world will actually unfold, *based on all available evidence*.

basis for action.

Nevertheless, current practice is certainly disappointing[11]. I suggest that there are persistent, systemic problems throughout aiding practice. (My experience bears principally on my specialty, decision analysis, though I doubt it is the worst offender). However, I am confident that addressing behavioral rather than technical issues can rectify this situation.

2.3 Essentials of useful aid

For an aiding effort to be successful, a number of absolutely essential requirements must be met (Brown 1989)

a. The product/output of the effort must be relevant to the decision.

b. All relevant knowledge available must be drawn upon.

c. Any model must specify appropriate inputs and connect them logically to output.

d. Input to the aid must faithfully reflect available knowledge.

e. Findings must be available in time to be used.

f. The D must accurately understand the meaning and implications of results.

g. The whole aiding process must be acceptable to the host institution.

If any one of these requirements fails to clear a critical threshold, the aiding effort will fail; and often does a number of reasons, including conflicted aider motivation.

3 MOTIVATIONAL HYPOTHESIS

While adequate development of QDA technology aiding looks promising, using it effectively is not assured. This section develops a causal hypothesis, a conceptual framework to explore its application, and an illustrative case.

3.1 A causal hypothesis

The role of motivation in aiding. Among controllable causes of aid shortcomings, I believe, the most serious are motivational, because they are both pervasive and resistant to change[12]. D himself may withhold critical information about his priorities from A. Economic or organizational pressures may divert an intermediary client. A has distinct

[11] However, our own old decision analysis text (Brown et al.1974) reflects an analytic technology that is still being taught, albeit in better texts (Clemen 1996), though I have since recanted many of the application prescriptions (Brown 1992).

[12] There are other influences on the usefulness of an aid that have nothing to do with A motivation, such as the availability of relevant information.

professional priorities. I will be concentrating on the latter, because of its importance and because the other motivations operate through him.

Motivational effects can, of course, be positive, such as pride in useful product, or desire to make the boss happy. As noted above, positive A motivation was a critical ingredient for aid success in WWII. The A in WWII had no conflict of interest. Both he and the Ds he served (e.g. military staregists), as a British citizens at mortal risk, were motivated to do what it took to defeat the Nazis. However, that overwhelming drive for As to produce realistic judgments receded after the War[13].

The central hypothesis. A's priorities commonly conflict with those of the decider they are supposed to aid, such that the aid they produce does not meet at least one essential requirement, and the aid is of no use to the decider. When A is not subject to an overriding motive to help the D make a sound choice, other A motivations come into play that work at cross-purposes. They reduce an aid's usefulness, even to the extent of distorting the D's judgment, leaving him worse off than before. This is commonly now the case.

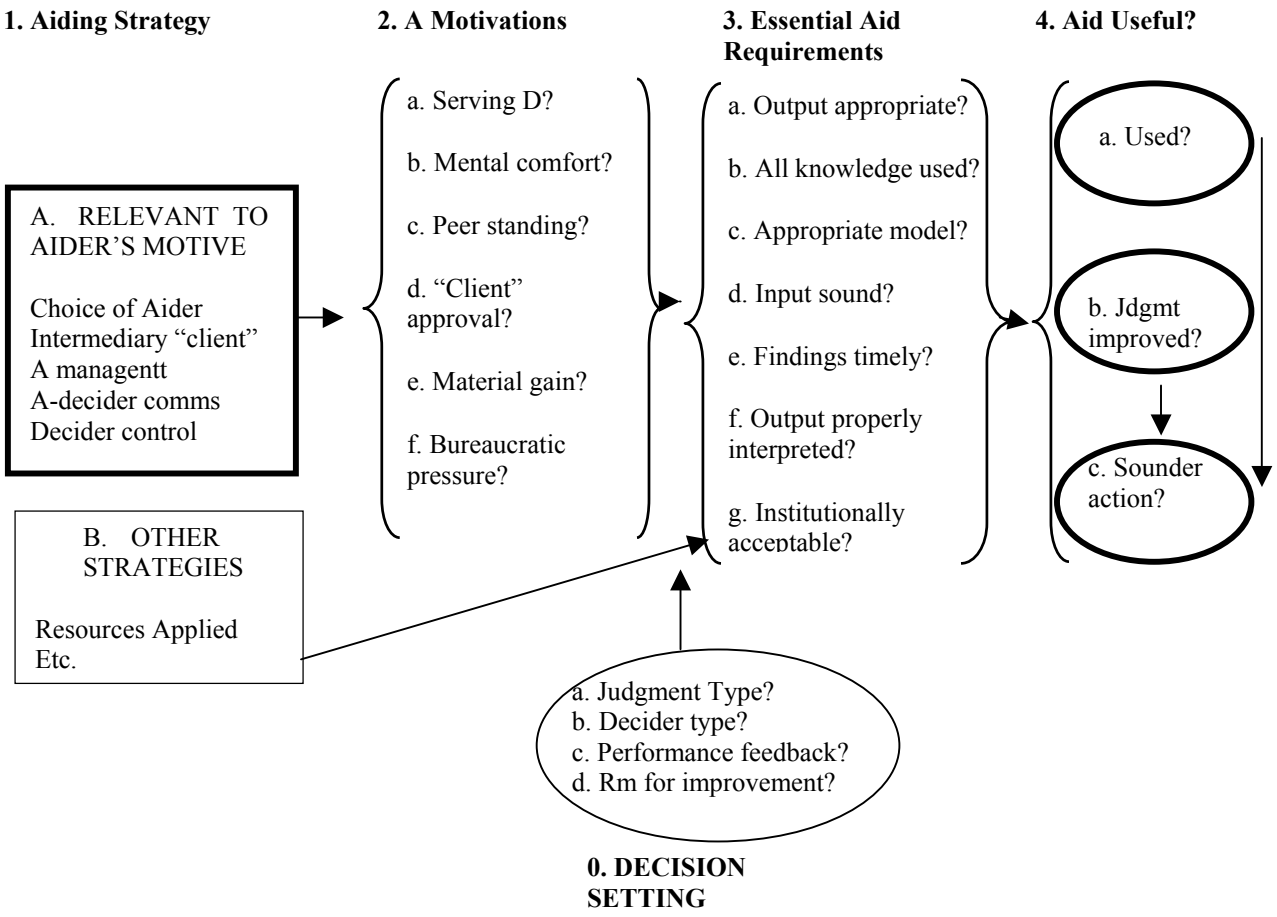
3.2 Framework for analyzing aid usefulness

Figure 1 presents a framework to explore systematically how an aid can distort D's judgment in a variety of circumstances. It shows the potential impact of aiding strategy (column 1) on A's motivations (col. 2), and the impact of that motivation on requirements of the resulting aid (col. 3) that determine its usefulness (col.4). The arrows connecting boxes show causal influences[14].

The top left box represents decision aiding strategy elements that typically influence aider motivation. They include working through an intermediary client, whose influence is on the "client approval" aider motivation in column 2. Strategy that does not influence A's motivations are summarized in the bottom left box[15]. The long arrow from it to the "aid requirements" column 3 represents all influences on essentials other than through aider motivation.

[13] A telling exception is election forecasting, where immediate feedback provides motivation and guidance to get it right—and usually it is got right (Florida 2000 debacle notwithstanding).
[14] Only the most significant hypothesized variables and influences are shown.
[15] Figure 1 is a simplified informal version of an "influence diagram" (Howard and Matheson 1983)

FIGURE 1. IMPACT OF AIDING STRATEGY ON A MOTIVATION AND AID USEFULNESS



Three attributes of a useful aid (col. 4) are: the aid is used; it improves a D's judgment; and/or it leads to sounder action. For an aid to have these attributes, it needs to satisfy a number of *essential* requirements (col. 3). Failure to satisfy critical requirements frequently stems from motivational pressures on A (col. 2). A's motivations in turn may be influenced by the aiding strategy (col. 1).

The decision task setting itself may mediate or exacerbate these influences (list 0 at bottom of figure). For example, take "performance feedback" (0c). If the setting is a company making short term investments the profitability of choices will be soon apparent, thereby rising the A's priority for usefulness in an aid (compared with a long term investment whose profitability may only become known when A expects to be long gone).

3.3 Illustrative case: reactor regulation

Reactor operators spend millions of dollars on the Probabilistic Risk Assessment (PRA) of reactor accidents that are mandated by the Nuclear Regulatory Commission (USNRC, 1990). This is an example of an aid that supports only one judgment in the decision process—assessing uncertainty—rather than directly indicating a final choice.

PRAs often produce assessments that disagree sharply with the judgment of seasoned regulators. In one case (noted earlier), PRA produced an overly optimistic safety evaluation of a reactor, which the regulator with decision responsibility found implausible, based on everything he knew. We explored with him how this might happen and concluded that he would be wise to rely on his own judgment, augmented, if possible, by selective contributions from the PRA. Accepting the PRA as it stood would be seriously misleading.

The following discussion recasts our informal deliberations into the more structured format of figure 1. Significant features of the aiding process (col.1) were related to its ultimate usefulness (col.4), via the impact of that process on the motivation of the PRA technicians (col. 2) and how that impaired usefulness of the PRA they produced.

3.3.1 Usefulness of aiding strategy

NRC's aiding strategy (col. 1), following common practice, included making the "regulatee", i.e. the reactor operator, responsible for assessing the safety of his own reactor (subject to regulator's concurrence). The regulatee spent several million dollars on PRA (Probabilistic Risk Assessment) as the aiding technology. The analytic work itself was contracted out to a company specializing in PRA, whose technical staff was largely engineers. There were thus three organizational steps of aiding service: D to regulatee, to PRA contractor, to A. Thus there were four parties with possibly different priorities.

The outcome of this exercise was that the regulator did not make any use of the resulting PRA, because he was not satisfied that it was an improvement over his unaided judgment—quite the opposite, in fact. We cannot automatically assume that the aiding strategy was responsible, but I maintain that it largely was.

3.3.2 Aid defects and their motivational causes

The immediate cause of the PRA's lack of usefulness (col. 4) appeared to be that it failed to meet certain critical requirements (col. 3) discussed below. These defects could, in turn, be traced largely to motivational impediments (col. 2), which were the direct consequence of the aiding strategy (col.1).

"Appropriate output" requirement (3a). The target risk assessment *should have been* for a severe accident from *any cause*. This PRA only addressed "internal risks" (like accidents due to hardware failures) and ignored "external risks" (like earthquakes). Even among internal risks, attention was limited to those that could be authoritatively documented, ignoring all others (like management failures) on the grounds that "we don't have the data" (to answer the right questions).

A large part of the cause was the As' mental comfort[16] (2b). The technical staff developing the aid was steeped in an engineering culture that made them uncomfortable making assessments that cannot be backed up with "hard facts" yielding "objective" probabilities. Engineers, making up most PRA technicians are trained to assess the "reliability" of stable processes, which a unique accident is not.

Moreover, A's client, the reactor operator has an obvious "material gain" motivation (2c) in having the reactor appear safe enough to avoid NRC penalties. He thus benefits from disregarding some of the risks. However, NRC bears responsibility for allowing the regulatee to assess his own risk—a clear conflict of interest.

"All knowledge used" requirement (3b). Even for those risks that *are* addressed, only *documentable* knowledge was used in the PRA (e.g. historical or experimental data); again stemming from engineers' discomfort (2b) with "soft" data. D, the regulator, in fact had much relevant knowledge beyond this. It included his personal observation that safety culture in the reactor was lacking, and that there had been a number of recent near misses at the plant and unfavorable evaluations by resident NRC inspectors. Though not definitive, it was knowledge he rightly wanted to take into account, and it was not in the PRA.

"Output understood" requirement (3e). The fact that only part of the risk was assessed (3a) would matter less if the regulator were made to understood the omission, so that he could informally factor in other knowledge[17]. The reactor operator's risk of economic penalty gave him, and consequently A, his PRA specialist, no incentive to make this clear (2e).

Impact of "decision setting" (list 0). The regulatee's "serving decider's interests" motivation (2a) was weakened by the absence of "success feedback" (Oc). The rarity of

[16] A technical group within NRC, whose motivations were comparable, prescribed much of the methodology.

[17] Misrepresenting its scope might be minimized if PRA were understood to stand for "Partial Reliability Analysis".

reactor accidents means that an unrealistically optimistic risk assessment—i.e. low accident probability—was unlikely to be revealed in time to embarrass its authors.

3.3.3 Strategy implications

This causal exploration of the indirect impact of aiding strategy on aid usefulness points to changes in that strategy. In particular, NRC might handle its own risk assessments through an A who is employed or engaged directly by D, the regulator, rather than enjoying the surely misplaced economy of leaving the responsibility to the regulatee.

4 AID ESSENTIALS OFTEN LACKING (3)

"The operation was a great success, but the patient died"

The above section sought to clarify and illustrate the motivational hypothesis. The present section demonstrates its more general applicability, with a variety of case anecdotes representing a full range of aid essentials that were lacking, i.e. fatal deficiencies (col. 3).

4.1 Inappropriate output (3a)

"Look for your keys in the shadows where you lost them, not under the street lamp where the light is better".

A valuable aid must first of all address the right problem, i.e. be designed to produce output relevant to the problem.

When to fire a torpedo. Navy high command felt that submarine commanders waited too long to fire torpedoes at enemy submarines in fleet exercises, running an unacceptable risk of being sunk first. A (myself) was asked to develop an aid to help the sub commander, as D, to choose when to fire. Our aid took into account D's probabilities of detection and counter-detection and the Navy's value tradeoff between "killing" and "being killed". Applying the aid to data from past exercises indicated that firing was indeed too delayed.

However, Ds revealed that their "non-optimal" action was not due to their poor decision-making, but to the fact that their priorities were different from the Navy's. Ds were evaluated by how accurately they located the enemy, and were not penalized if they were "sunk". The commanders already rationally served their own interests, and needed no aid. I had addressed a non-problem[18].

4.2 Knowledge overlooked (3b)

There are several variants of critically overlooked knowledge.

[18] However, it did uncover a real decision problem for the *Navy* higher command, as D. How could Navy align a commander's priorities with the Navy's, so that he would act as our aid prescribed? (Clearly not a job for QDA).

4.2.1 Plural vs. mono-evaluation

Choices and other judgments can be approached in different ways that draw on different types of knowledge. What I call "mono-evaluation" is the common decision aiding practice of basing a target judgment on a single model or approach, thus ignoring relevant knowledge that alternative approaches could tap into. (See Appendix A.) "Plural evaluation", on the other hand, addresses a given judgment several ways and combines findings, thus accessing more knowledge[19]. A simple but important example of plural evaluation is to confront aid findings with "expert opinion" (neglected with serious consequences in the above nuclear safety case). Another is to compare findings with observable reality (external validation).

However, until recently mono-evaluation has been the dominant practice, with some significant exceptions[20]. As tend to shy away from plural evaluation, no doubt because it risks embarrassing discrepancies between findings—the possibility of which is exactly what makes plural evaluation so valuable[21]. As preserve confidence in their findings by restricting themselves to a single approach, with no risk of inconsistency. Without plural evaluation, a misleading model easily passes unchallenged (see parts depot case below).

4.2.2 Undocumented knowledge

"If I were going there, I wouldn't start from here"

As illustrated in the nuclear safety case above, engineering and scientific cultures typically refuse to draw on knowledge that is not properly documented and validated.

On uncertainties where both documented and undocumented data exists (as often on machine reliability), the consequence of ignoring the latter is a judgment that is based on incomplete knowledge. However, on issues where the *only* knowledge is undocumented (as often with human error), the issue is more serious. A must either acknowledges that he cannot address the uncertainty, or fall back on some arbitrary assumption (such as a zero probability). Such an approach would likely have evaluated the risk at Chernobyl as negligible, since the major problem was undocumented managerial failure.

4.2.3 Unknown unknowns

"There are more things in heaven and earth, Horatio, than are dream'd of in your philosophy." (Shakespeare)

---

[19] Although formal procedures for doing plural evaluation have been developed (Brown 1976, Lindley et al. 1979) most of the benefit can be obtained informally. Top-down modeling as commonly practiced is a special case of mono-evaluation, in that at the end of the process judgment is based only on the most disaggregated model.

[20] Notably forecasting (Winkler and Makridakis 1983).

[21] The discrepancies may pass unnoticed. Sex surveys typically estimate that men have more heterosexual partners over their lifetimes than women. Barring some far-fetched assumptions, arithmetic says they average should be the same, i.e. they are alternative estimates of the same phenomenon. Acknowledging this, however, would force surveyors to concede that at least one estimate was wrong—presumably because the men overstated and/or the women understate their romantic experience.

A major type of knowledge often overlooked by QDA is the awareness that things may happen that have not been identified in advance—the "unknown unknowns" problem. .

Predicting human intrusion into waste dumps. To comply with regulation on permissible radioactive emissions from a nuclear dumpsite over the next 10,000 years, DOE considered risk from "human intrusion". PRA technicians identified only one plausible intrusion scenario: mining for minerals near the repository, to which they gave a plausible probability of 0.1%. However, this 0.1% was reported as the probability of *any* human intrusion. Any other human intrusion was effectively treated as impossible. The unreasonableness of this position is demonstrated in Appendix B.

Forecasting new demand. Historically, demand for new commercial products has been at least double what had been predicted ten years earlier on the basis of all *identified* applications. A 1985 government study predicted steadily *declining* demand for space on commercial satellites, simply because *known* prospects were fewer the further ahead one looked. Acknowledging that phenomenon realistically would have yielded less pessimistic and more realistic forecasts of commercial space demand (which has in fact increased greatly since then).

4.2.4 Criteria omitted.

The most critical knowledge commonly overlooked is some non-obvious *criterion* about which the D is in fact most concerned (as demonstrated in Appendix A).

D's motivation does *not* appear to be the cause in any direct way; rather A's lack of value elicitation skills. There are a couple of paradoxical exceptions.

D may be embarrassed to disclose some sensitive objective, and as a result his judgment is distorted *even for his own purposes.* A company executive may wish to line his pockets at the expense of shareholders (as in the Enron case). A civil "servant" may sacrifice public service in order to avoid bureaucratic hassle[22]. However, it is D misleading A, not the other way round.

A may feel under moral pressure to serve the public, rather than D. I was helping a congressman decide whether to support a community anti-crime bill. He was concerned that it would cost him political support from the Association of Chiefs of Police. However, he was reluctant to have me include this criterion overtly in my aid, though it would tip the balance against the bill. To ease my conscience, I gave him the "public spirited" evaluation, without the politics. However, I also provided him with a parametric version of the aid that allowed him to secretly input his personal priorities. He voted against the bill.

[22] A nuclear regulator told me that, when deciding what reactor safety measures to require, it would be embarrassing for him to acknowledge publicly in a formal model that he took into account political "hullabaloo" (of the type caused by the Three Mile Island accident), as well as legitimate health effects (apparently negligible at TMI). His A did not take this into account and so understated risk (as D valued it)

4.3 Inappropriate model (3c)

This requirement is different from the others. It currently tends to *dominate* analysis, for reasons of professional priority, to the detriment of the other requirements. This can have disastrous results (see parts depot example below). There are two related sub-requirements. Output must follow logically from input. And the inputs it specifies must be appropriate, in the sense that they can be derived soundly enough to imply output that improves on unaided judgment. Model complexity does not necessarily assure this (Brown 1991).

In fact, some sacrifice of logical soundness may be sensible. A logically flawed QDA may outperform more rigorous analysis that fails *other* aid requirements. For example, AHP (Analytic Hierarchy Process) appears logically less rigorous than MUA (multiattribute utility analysis) (Belton and Gear, 1983; Winkler et al 1990.). However, AHP appears to be used more widely than MUA, possibly because it performs better on *other* aid requirements, such as sound input (3d) or institutional acceptability (3g). If Ds won't use MUA, or not successfully and AHP it gets decisions closer to ideally rational choice than using no aid at all, AHP would appear to the method of choice. Nevertheless, aids do need to clear a soundness threshold, and they usually do.

4.4 Unsound input (3d)

"Garbage in, garbage out."

If the best knowledge available is poor, some input "garbage" may be unavoidable, whether aided by QDA or not[23]. However, an aid may introduce unnecessary input garbage, if that knowledge is poorly elicited or derived.

"Don't spoil the ship for a ha'porth of tar"

Whether to close parts depots. Ford UK felt they had too many car-parts depots in the London area. A prestigious university OR group developed a transportation-optimizing model that indicated that only four of the seven were needed. Ford closed three depots, with disastrous results. The four remaining depots proved completely inadequate for the demand. It turned out that depot capacity, a critical input, had been grossly overestimated. As had casually calculated usable capacity as height times length times width—without taking dead space into account.

This input error could easily have been avoided, by checking it for plausibility with any Ford stock controller (a simple plural evaluation). It was not, due to lack of balanced attention to essential aid requirements. The model structure was no doubt fine (3c), but all available knowledge was not used (3b). The As presumably considered "serving D's interests" (2a) too low a priority to make even a small effort to assure it. Indulging their technical expertise, doing what they know best and furthering their professional reputation (2b) may have been their overriding priority. The D's strategic mistake was to

[23] QM has the potential to avoid adding "garbage in-between" to the "garbage in", by making sure that the garbage is at least "processed" properly, i.e. subjected to logical discipline.

engage dedicated academics (1b).

This is an old case study (late 50s), but it shows in a particularly sharp way pitfalls that persist to this day. What is dated about this case is that clients then had so much misplaced trust in QDA that they were prepared to act on it. Such imprudence would be rare today (but see the nuclear short-listing case in 4.6.1 below).

Probability of geologic events. As part of the nuclear waste-siting project discussed earlier, a model of radioactive release was developed, with probabilities of geologic conditions supplied by experienced geologists. When, as peer reviewer of the exercise, I asked an expert how he did it, he said, "When asked for a probability I always say ten per cent". He was tongue-in-cheek, of course, but he was clearly casting doubt on how validly his assessments reflected everything he knew. His and other experts' input may have been ill considered because input quality was not high priority in the study. Time and effort may instead have been devoted disproportionately to model quality (3c).

4.5 Untimely findings (3e)

"The best is the enemy of the good"
"I want it Tuesday rather than right!"

Professional perfectionism often delays aid until is no longer useful. A technician may hold back incomplete work, even if it is adequate for policy purposes, because it lays him open to damaging peer criticism (2c)—or at least to personal dissatisfaction (2b).

Evaluating Clean Air Legislation. In 1990, a Democratic Congress asked EPA to evaluate whether the 1970 Clean Air Act had been worth its cost, as input to future legislation. My task was to integrate a number of major chemical, economic and environmental modeling studies into an aggregate model of total regulation impact (Brown 1991). The researchers involved resisted providing any results until these were definitive enough to survive peer scrutiny (2c), even though they were perfectly adequate for practical purposes. Months later, and well before this threshold was cleared, Republicans took control of Congress and dropped the project, before it could inform policy.

The strategic aiding error was allowing the scientific As the discretion to do what research they pleased and when to report it.

4.6 Misinterpreted Output (3f)

A further problem in this Clean Air case was that the same researchers cast their results in a form that suited them, rather than a form adapted to project needs (3a). I.e. their outputs did not coincide with inputs called for by my policy oriented integrating model.

A moderate mismatch between aid output and problem needs may be acceptable, if the gap is acknowledged, so that the downstream modeler (myself) can adjust is input

accordingly[24]. However, the modeler may be less technically qualified than the primary researchers to make such adjustments.

Short-listing nuclear waste sites. An award-winning decision analysis ranked candidate sites for a nuclear waste repository, based on preliminary evidence. DOE staff published a recommendation to study intensively the top three sites -- two in salt and one in basalt (DOE 1987). The Energy Secretary disagreed and substituted a third rock medium, known as tuff, in place of a salt site. Critics accused the administration of disregarding authoritative decision science for improper political purposes.

Be that as it may, I believe the Secretary's decision to include all three rock types was quite compatible with sound decision analysis. The DOE staff had misinterpreted *ranking* as *short-listing*, which would call for "portfolio diversity"[25]. If a salt medium appears best under current knowledge, it is not surprising that two of the top ranked sites should be in salt. However, since it was not yet clear which medium was best, it made sense to retain all three to study further. At the DOE staff head's request, I expressed this argument briefly in decision analysis terms that was compatible with the study (Brown 1987b)[26].

Alternatively, the wrong problem may have been addressed by the aid (requirement 3a), as a result of not identifying short listing with portfolio selection. This could be attributed to lack of timely communication between D and As, itself possibly brought on by reluctance of both parties to share delicate considerations with the other (motivation 2b).

In either case, I suggest the problem could have been avoided with more "decider participation" (1a) in the analytic process. The As or DOE staff could have reviewed with D, the Secretary, early tentative findings of the study (which already pointed to the same three top ranked sites). He would likely have balked (for the portfolio reasons) and the study could have been redirected. In fact, in this case the decision proved worse than useless, because of political embarrassment it caused.

Energy conservation policy. During the 70s oil crisis, DOE's predecessor agency, FEA, considered energy conservation measures (as an alternative to increasing energy imports or production). A key judgment involved in the decision was assessing the current level of household conservation practice. A high level would support the argument that there was little scope for further conservation and no incentives were called for. Survey sampling was an aid aimed at enhancing a component judgment in deciding.

[24] In physics, the acceleration model of an object falling in vacuum is useful for predictions under air resistance in a normal non-vacuum (unless the mismatch is too great, like for a falling feather).
[25] This was later pointed out by a study author (Keeney 1987)
[26] Essentially, it said that if, after further study, one salt site should appear unacceptable, the conditional probability of another salt site appearing acceptable maybe less than for a site in another medium.

In a survey of 1000 households, a well-known pollster[27] reported to FEA policy staff that 61% already lowered thermostats to save energy, which corresponded to the fraction so claiming in the sample. However, FEA staff was skeptical and asked us to reevaluate these findings. We conducted an intensive on-site calibration survey of a small subset of original respondents, and found that about half of those claiming to lower their thermostats in fact didn't. Policy staff lowered their estimate from 61% to 30% and acted accordingly (Brown et al.1977),

Survey biases in general. The above case is an example of a common phenomenon with sample surveys. When no "objective", i.e. judgment-free, way to assess measurement error and bias (or other non-random errors) is apparent, the latter are treated as zero, either for reasons of tractability and therefore mental comfort (2b), or to promote client confidence in survey findings (2d,e).[28]

4.7 Institutional unacceptability (3g)

Decision aid may fail because the aiding process is unacceptable to the host institution. For example, policy makers often resist having their decisions subjected to reviewable rationale for organizational reasons, such as turf, corporate culture and loss of discretion (Porter 1989; Brown 1970). William Ruckelshaus (previously Secretary of the Environment) told me he was opposed to Congress requiring the Executive Branch to justify decisions with QDA, on the grounds that its transparency would increase Congress's ability to interfere.

Aid may be institutionally unacceptable because the aiding process disturbs an organization's power structure (Brown 1970). For example, it may devalue traditional decision skills and therefore the managers whose strength these have been. The managers may feel less threatened if A reports to them. The aid then represents enrichment of their capabilities rather than competition.

A's priorities are indirectly responsible for the institutional resistance problem, except indirectly, because any serious aid deficiencies (3) that they cause reduce confidence in *any* aid. For example, in the above unfortunate nuclear waste siting case, the DOE manager with oversight responsibility for the original study told me that he would never again use QDA on a high visibility decision.

5 SOURCES OF CONFLICTING MOTIVATIONS (2)

Understanding the sources of A's conflicting priorities may give guidance on what can be done to reorder them, to mitigate their effects, or to choose a different A.

[27] The Gallup Organization.

[28] Incidentally, treating random error as the only measure of estimate accuracy encourages the large sloppy samples (with little random error) rather than smaller, carefully measured ones (which may be more accurate overall) (Brown 1976).

5.1 Serving D's interests (2a)

This is the motivation D would *like* to have dominated. It can be done by eliminating other motivations (see below) and/or increasing this one, say, by having A report directly to D and using people without academic aspirations. However, this may come at the cost of the technical expertise that an independent high-powered support group could provide (Brown 1970).

5.2 Mental comfort (2b).

"To a man with a hammer everything is a nail"

The most important "pain avoidance" practice for an A may be "doing what you know best" i.e. exercising his particular expertise and avoiding anything that involves unfamiliar thinking (no matter how critical this may be to successful aid). A mathematician may only address the quantitative modeling aspect of an aid (3c) and completely disregard any behavioral issues (3f and 3g)[29]. QDA practitioners tend to favor definitive answers to minor problems over tentative answers to major problems—even when the latter is more useful.

"Good policy analysis needs bad science" (Granger Morgan 1978). I would add to this "Sophisticated modeling may produce naïve decision aids". The A is often steeped in the professional culture of a scientist who has been trained to search for eternal truths and will not sign off on findings that he cannot authoritatively document. Other "mental comforts" include:

- Hassle avoidance
- Intolerance of ambiguity;
- Embarrassment over admitting uncertainty;
- Analytic tractability.

5.3 Peer standing (2c).

A may seek approval in the academic community, especially if he is university faculty. So he is led to produce quantitative models that are internally coherent, technically sophisticated and free of the risk of embarrassing flaws, but do not promise success in the real world.[30]

5.4 Client approval (2d)

A client interposed between D and A may have objectives that conflict with D's. British decision scientist Stephen Watson[31] notes, "Quantitative research is often commissioned by those who have no responsibility for implementing any resulting recommendations,

---

[29] March and Shapira (1982) cite an analogous case where company directors spent most of a board meeting discussing where to put a bicycle shed and five minutes on whether to invest $500,000 in a new energy technology (which they were less comfortable talking about).

[30] I painfully remember regretting publishing an innovative probability procedure when a fatal flaw in it were pointed out at a professional meeting I was at (Keefer 19xx)

[31] Personal communication to author, 2000.

nor any stake in their success. They just have to be seen to be 'doing something'". The need to keep the client happy and retain profitable business obliges the A to adopt his objectives. A remedy is in the selection or motivation of that client, second order measures that stretch the scope of this paper.

Clients are particularly prone to conflicting motivations when some outside body imposes a research requirement. For example where a regulatee is required to evaluate his own performance, supposedly as an aid to the regulator D, where companies hire their own auditors (as in the notorious Enron case), or legal disputants hire their own expert witnesses.

Aviation safety An FAA official explained to me why he had awarded a risk analysis contract to someone else. "Your proposal was better, but his was cheaper. Executive Order 12291 obliges us to do a risk assessment before we make a major change in flight procedures. No reason to spend more than we have to." Part of the problem is misplaced economy by the ultimate D, who controls the purse strings, the Administration here. Requiring QDA, without controlling for quality, motivates A to make cost his first priority and sacrifice usefulness.

5.5 Material gain (2e)

The "client approval" motive partly subsumes "material gain" if it promotes A's career or profitable follow-on business. Material gain if A is a consultant may also encourage him cut costly aid improvements.

5.6 Bureaucratic pressure (2f)

This is again similar to the "client approval" motivation, but tends to be more diffuse. For example, a British machine tool maker had A evaluate how to reduce wastage of an expensive cutting material. However, the hidden, but internally recognized, company agenda was to persuade its holding company that they could do nothing because the fault was with the material supplier, another company in the group. A understood that he was expected to come up with the wrong answer``.

6 CONCLUSIONS

6.1 Findings

In this paper, I have focused attention in two directions. The interests to be served are those of D (whose priorities may not be those of the organization or community that he supposed to serve). The object is to make *his* judgment more realistic[32], for whatever purposes he may have. In addition, the only source of an aid's failure to improve the D's judgment that I have address directly is conflicting *A* priorities (and only indirectly, those of an intermediary "client").

32

I have argued that any failures in the application of QDA (and qualitative aids, for that matter) have little to do with inherent flaws in the analytic technology. Decision practice needs improving, and QDA of various kinds can do the improving. Any defects are often technically easy to remedy. Failures are caused largely by motivations that distort aiding efforts. Needed technical development will come in time, if not the motivation to use it. If the aid usefulness problem were only a matter of immaturity in aiding art, normal technical evolution would no doubt cure it.

6.2 Remedial strategies for the decider (1)

"Caveat emptor!"

Sections 4 and 5 discussed aid requirements, A's motivations and D's aiding initiatives that might influence them. This section attempts pull the strands together into broader strategic suggestions.

Note, however, that total aiding strategy includes more than what directly affects A's motivation, such as the resources made available to A. However important it is to assure appropriate motivations, lack of resources may prevent an aid meeting critical requirements, regardless of A's motivation. Moreover, the usefulness of aids that do impact motivation depends on more than that impact (e.g. the cost of the aid and the room for decision improvement).

I have argued that *what* goes wrong in aiding is mainly unbalanced attention to all the essentials of a useful aid. My motivational explanation of *why* it goes wrong suggests is that it may take some kind of decider-sponsored revolution to make balance happen.

6.2.1 Choosing the right aider

D could engage an A whose motivation is already appropriate and shares D's goals. A for-profit consultant, hungry for follow-on business from D and familiar with the practicalities of the business, may be better than an academic looking to tenure in an OR department, however impressive his technical credentials. On the other hand the consultant's motive to cut costs may be even more harmful.

The aiding team should be multi-disciplinary (to cover all the critical bases of successful aid). A social scientist, for example, can assure that cognitive issues are considered. If the As are all mathematicians, without cognitive qualifications, they may be competent to produce logically sound models but not to elicit accurate input (3d).

Organizational and other human factors have been shown to represent about half nuclear accident risk, but account for 5% of NRC's research effort. When a budget crunch forced economies in the NRC Office of Research, it was this 5% that was reduced. Budgets go to powerful players. At NRC these have been the engineers, for whom aid usefulness to Ds or society may be less important (or less well understood) than solidarity with their own kind.

6.2.2 Motivating the chosen A

"He who pays the piper calls the tune"

No amount of exhortation on how to "do it right" will get very far, if A is not motivated to do it right. Having A report directly to D may be the best way to ensure that A's priority is to serve D. However, organizations often set up a specialized central aiding team (e.g. OR department) to serve all other departments. This has the advantage that the cost of a high-grade resource can be spread throughout the organization. However, their careers are less dependent on the goodwill of those they aid, and are more subject to a technician culture (Lorsch and Lawrence 1968; Brown 1970).

6.2.3 Decider-Aider communication

Ds should insist on maintaining thorough, continuous and unreserved communication with their QDA helpers. If the D is not involved in the day-to-day development or implementation of an aid, he may be in no position to oblige the A to do what is needed (2a)—as opposed to what A is comfortable doing (2b). The result may be institutionally unacceptable (3g) and therefore not used (4a).

Monitoring aid progress. It helps for the D to stay on top of the analysis and to check out the plausibility of findings against the judgment of experienced staff well before acting. At a minimum, the client can require that preliminary findings be reported early in the analysis to check that the right problems is being addressed and that any conflict with D judgment be explained by reasons *other than* the provisional nature of the analysis. (This could have avoided serious embarrassment in the above nuclear waste short-listing case). I usually find that a plausibility check with my client early in a project uncovers enough wrong for me to redirect it (or to regret later that I had not).

6.2.4 Directing the Aiding

"Decision aiding is too important to be left to the decision aiders."

Once A has been chosen, it may be very difficult to reorder his priorities so that "serving D" or "client approval" are near the top, if they are not at the outset. The cultural and cognitive proclivities of professional As capable of doing the job may be inconveniently stable. Aid users might then have to be more intrusive in the design and use of QDA aids if they are to get significant benefit. Major intervention by the D or beneficiary will be needed to check that an aiding effort is on track, and to put pressure where needed. However, to take active charge of the aiding process, D must be competent to specify what the aid does, check its substance and direct its use, which is a non-trivial condition. Moreover, the control must be highly selective, to avoid the classic dangers of micro-management. In particular, D should limit his intervention to aiding actions that he can confidently attribute to unsatisfactory motivations

6.2.5 Intermediary "client"

"Don't have the fox guard the chickens".

Since the focus of this paper is on the *aider's* motivations, I am only concerned with his "client's" motivation—important though that is to aid usefulness—to the extent it is transmitted to A, via "client approval".

One such client motivation has a critical impact on A. Any third party put in the line of command from D to A, e.g. as A's client, should not have a vested interest in any of the options considered. For example, a regulator should not delegate risk assessment to the companies it regulates (as was done in the reactor safety example). By all means let the regulatees bear the cost, but have D take charge of the aiding effort itself (including hiring, and therefore motivating, specialists).

6.3 *Validity of conclusions.*

The argument I am putting forward is based on 40 years of experience as an aider and aid advisor to executives, as well as on broader surveys of aid users (Brown 1970, Brown and Ulvila 1982, Brown 1987). Although this "data base" covers hundreds of decision problems throughout government and business, it is essentially anecdotal and the interpretations highly subjective (but hopefully well-informed).

Nevertheless, my observations are reasonably consistent over a wide variety of decision problems and settings, of which those cited in this paper are only a small sample. My case file covers executive, legislative and judicial decisions in virtually all the major government agencies and domains, as well as a more limited number of marketing, manufacturing, product policy and investment decisions in business. This consistency encourages me to believe that the hypotheses based on this experience are worth at least checking out

6.4 Work to be done

The main task, I suggest, would be to extend and validate these findings, by making the sampling deliberate (rather than opportunistic) and, much more important and difficult, to devise and apply less subjective methods to characterize and measure cause and effect in each chase.

A promising area of aiding research, with enough scientific challenge to interest the best minds, is *measuring* the usefulness of decisions and decision processes (Brown, 1994). This could lend authority to findings that so far depend on my undocumented judgment. The topic has been largely neglected, partly, I suspect, for another motivational reason: it could challenge the validity of other QDA work.

6.5 Parting thought

There is a certain irony in a practitioner and teacher of QDA appearing to caution potential customers *against* their use--at least in certain cases. On the contrary, I am cautioning customers *in* their use. Publicizing QDA pitfalls may help improve practice, enhance the field's reputation, expand its effective reach and increase demand. Weeding out malpractice in decision aiding, as in medicine, can only help the field. But it will take the Ds themselves learning enough about these pitfalls and their cures to keep us aiders honest—and useful.

ACKNOWLEDGMENTS.
I am grateful to a number of colleagues for helpful suggestions and comments, including, Stephen Watson, Jonathan Baron, Baruch Fischhoff, and especially John Lathrop and two anonymous referees.

REFERENCES

Allison, G. Essence of Decision: Explaining the Cuban Missile Crisis. Boston: Little Brown, 1971.

Baron J. Judgment Misguided: Intuition and Error in Public Decision Making. New York: OUP. 1998.

Belton V and Gear A. Evaluation of AHP. Omega, 11, 228-230, 1983

Brown, R.V. Do managers find decision theory useful? *Harvard Business Review*, May-June 1970, 78-79.

Brown, R.V. Rationality and the Resolution of Incoherence. Working paper. Department of Statistics, University College, London. November 1976.

Brown, R.V. Republican National Committee vs. Federal Election Commission: Deposition of Dr. Rex V. Brown in the U.S. District Court Southern District of New York. May 1979.

Brown, R.V. Characterizing a nuclear waste site: A logical framework for allocating resources. Proceedings 1988 Annual Meeting of the Decision Sciences Institute, November 1988.

Brown, R.V. Toward a prescriptive science and technology of decision aiding. Annals of Operations Research, Volume on Choice Under Uncertainty, 1989, 19, 467-483.

Brown, R.V. Decision analytic tools in government. In Karen B. Levitan (Ed.), *Government Infostructures*. Westport, CT: Greenwood Press, 1987a.

Brown RV. Decision analysis validation of selected characterization sites. Memorandum to Ben Rusche, head of DOE Office of Civilian Radioactive Waste Management. Sept. 1987b.

Brown, R.V. Evaluating the Cost-Benefit of the Clean Air Act Realistically Using Hierarchically Partitioned Assessment (HPA). Presented to the Society for Risk Analysis national conference, Baltimore. December 1991.

Brown, R.V. The state of the art of decision analysis: A personal perspective. *Interfaces*, **22**, 6 Nov-Dec 1992, 5-14.

Brown, R.V. The Role of Statistical Decision Theory in Decision Aiding: Measuring Decision Effectiveness in the Light of Outcomes. In Freeman PR, Smith AFM (Eds). Aspects of Uncertainty: A Tribute to Dennis Lindley. Wiley, 1994.

Brown RV Predicting the decision aiding value of research. Working paper. School of Public Policy. GMU. Feb 2000.

Brown RV. Choice and Uncertainty in Environmental Risk Management. In G. Gorbaasi Encyclopedia of Life Support Systems. UNESCO. In press.

Brown RV. Logic and Motivation in Risk Research: a Nuclear Waste Test Case. Risk Analysis. Forthcoming a.

Brown RV. Tools of Rational Choice. Wiley. Forthcoming b.

Brown RV, Campbell VN and Repici RJ. Analysis of residential fuel conservation behavior. McLean, VA: Decisions and Designs, Inc., October 1977.

Brown, R.V., and Ulvila, J.W. Does a reactor need a safety backfit? Case study on communicating decision and risk analysis information to managers. *Risk Analysis*, 1988, *8*(2), 271-282.

Brown RV, Kahr AS and Peterson CR. Decision Analysis for the Manager. NY NY: Holt. 1974

Clemen R. Making Hard Decisions. Belmont, Ca: Duxbury Press. 1996

Computers and Operations Research Special issue: "Operations research and the rude intrusions of the real world." 18,7,1991.

DOE, A multiattribute utility analysis of sites nominated for characterization for the first radioactive waste repository. U.S. Department of Energy, 1987.

Flanders NF, Brown RV, Andre'eva E. Larichev OI, Justifying Public Decisions in Arctic Oil and Gas Development: American and Russian Approaches. *Arctic*. Vol. 51, no. 3, p. 262-279. September 1998.

Grayson CJ. Decisions Under Uncertainty. Harvard Business School, 1960.

Grayson, C.J. Management science and business practice. Harvard Business Review, July-August, 1973.

Howard RA. Heathens, heretics and cults: the religious spectrum of decision analysis. Interfaces, 22, 6 Nov-Dec 1992.

Howard, R.A. and Matheson, J.E. Influence diagrams, in R.A. Howard and J.E. Matheson (eds.) *Readings on the Principles and Applications of Decision Analysis*, Vol. II. Menlo Park, CA: Strategic Decisions Group, 1983.

House PW. Rush to Policy: analytic techniques in public sector decision-making. New Brunswick: Transaction Book, 1988.

Keeney RL. An analysis of the Portfolio of Sites to Characterize a Nuclear Repository. Risk Analysis 7. 195-218. 1987.

Keeney RL and von Winterfeldt D. Managing Nuclear Waste from Power Reactors. Risk Analysis. 1994.

Lorsch JW and Lawrence PR. Integration and Differentiation in Industrial Organizations. Harvard Business School. 1968.

March, J.G., and Shapira, Z. Behavioral decision theory and organizational decision theory. In Ungson, G.R., and Braunstein, D.N. (Eds.), Decision-making: An interdisciplinary inquiry. Kent Publishing Co., 1982.

Majone G. Anatomy of Pitfalls. In Majone G and Quade ES, Eds. Pitfalls of Analysis. New York: Wiley 1980.

Moore PG. Technique vs. Judgment in Decision-Making. Organizational Dynamics. Autumn 1973. PP 69-80

Morgan MG. Bad science and good policy analysis. Science. Vol. 201.p. 971. Sept. 15 1978.

NUREG-0880 *Safety goals for nuclear power plants: A discussion paper.* Washington, DC: U.S. Nuclear Regulatory Commission, February 1982.

Politser P "Do medical decision analyses' biggest gains come from the smallest decision trees?" JBDM, 4, 139-141, 1991.

Porter, H. Profile: A Decision Analyst Reflects on His Work. Chance, 2, No. 4, 38-40, 1989.

Quade, E. Pitfalls in Formulation and Modeling. In Majone G and Quade ES, Eds. Pitfalls of Analysis. New York: Wiley 1980.

Simon, H.A. Report of National Academy of Sciences Panel on Research needs for decision making. In Impact and potential of decision research on decision aiding. Eds. Holt, V. and Tolcott, M.A., American Psychological Association, 1988.

Strauch RE. A critical look at quantitative methodology. Policy Analysis 2. Winter

1976. pp 121-144.

Thomas H. "Implementing Decision Analysis". In Ira Horowitz (Ed.), *Organization and Decision Theory*. Kluwer-Nijhof Academic Publishers, 1989

USNRC, Severe accident risks: an assessment of five US nuclear power plants. NUREG-1150. Washington, DC, 1990NRC

von Winterfeldt, D. and Edwards, W. *Decision analysis and behavioral research.* New York: Cambridge University Press, 1986.

Winkler R, Dyer J, Saaty T and Harker D. Special issue on AHP. Management Science, 36, 247-275, 1990.

Winkler, R.L., and Makridakis, S. The combination of forecasts. Journal of Royal Statistical Society, 1983, 146, 150-157.

APPENDIX
HYPOTHETICAL DEMONSTRATIONS OF METHODOLOGICAL ARGUMENT

A. Importance of plural evaluation

The following thought experiment[33], demonstrates the potentially critical importance of "plural evaluation" (Brown 1976). A famous athlete has been found guilty of murdering his wife and her friend, after many months of riveting media coverage of the trial. As a decision aid to sentencing, the judge constructs a benefit-cost model to evaluate the consequences of the crime.

Benefit = 100M TV viewers @ \$10 average entertainment value to viewers
= \$1 Billion value[34]
Cost = 2 lives @ \$20 Million (consistent with some regulations[35])
= \$40 Million

The judge notes that this represents more than a 20:1 benefit:cost ratio. He might conclude that the murderer has performed a valuable public service (and so should be commended), were it not for his intuitive conviction that the ratio should be negative. He thinks hard about what must be wrong with the model, realizes that it disregards the precedent effect of society condoning murder, and corrects the model.

In effect he has used an informal form of plural evaluation, where the results of a model are tested against intuition. In real aiding cases, intuitive evaluation is not so convincing (which would make aid unnecessary).

B. Consequence of ignoring unknown unknowns.

The following fanciful analogy may help appreciate this issue. 10,000 years ago an Indian tribe was deciding whether to bury its ancestors' sacred bones 20 feet below their island home. Tribal regulation required them to assess the probability of human intrusion over the next 10,000 years. Their PRA high priest could think of only one plausible scenario: desecration of the site by a mainland tribe, whose probability he assessed at 0.1%. He reported a 0.1% probability of any human intrusion over the next 10,000 years. The name of the island was Manhattan…

The PRA high priest did not, of course, have our hindsight about perils from subways and such. However, it would be reasonable for him to judge it "as likely as not" that humans would do *something or other* to disturb their sacred bones over the next 10,000 years. If he were required to estimate *some* probability, it would certainly be higher than 0.1% (perhaps within a range as large as 10-90% range). A parallel argument can be made in the above nuclear waste siting case. I suggested that any realistic assessment might be somewhere in the range 10-90%. Even this could be wildly out, but if *some* number must be used in the overall risk assessment, it should certainly be more than 0.1%.

---

[33] Based on the celebrated 1997 OJ Simpson trial.
[34] Contingent valuation not actual flow of money.
[35] E.g. Nuclear safety according to regulatory guideline NUREG-0880 (1982)

About the author

Rex Brown has spent 40 years alternating policy consulting with research. He was chairman of Decision Science Consortium, Inc., where he advised senior government and business executives on decision aiding methods. He taught decision analysis at Harvard Business School and the London School of Economics and is shortly publishing a textbook on tools of rational choice. He was awarded the 1967 British Institute of Statisticians Prize in Applied Statistics.