Judgment and Decision Making, Vol. 1, No. 2, November 2006, pp. 162–173

Making decision research useful — not just rewarding

Rex V. Brown*
School of Public Policy
George Mason University

Abstract

An experienced decision aider reflects on how misaligned priorities produce decision research that is less useful than it could be. Scientific interest and professional standing may motivate researchers — and their funders and publishers — more powerfully than concern to help people make better decisions.

Keywords: decision analysis, decision aids.

"Doing science is like making love. Some good may come of it, but that's not why we do it." (Richard Feynman)

1 Introduction

40 years ago many of us thought we were on the brink of a new era: thanks to emerging decision analysis tools, we could look forward to a brave new world, where we would no longer make foolish mistakes that ruin our lives. So far, nothing like that is remotely in sight, and I don't expect it ever will be.

However, I still believe that decision aiding, and prescriptive decision analysis in particular, *can* become a major force for good in the world; but it will take an upheaval in how decision tools are fashioned and how they are used. It won't come about spontaneously, because the root problem is not technique, but motivation, which is notoriously difficult to correct.

Today, I want to talk about how we can make decision aiding more successful, by making usefulness a top priority in the decision aiding community and among decision researchers in particular. A lot of important decision research is being done, but not very much of it is helping decision aiding to get used and be useful. I think that can be turned around, but it won't be easy.

2 The problem

2.1 Decision aiding in the doldrums

People make terrible mistakes all the time. We marry the wrong person, our government takes misguided military action, and we pay for it dearly. Human welfare has greatly suffered through poor decisions.

For half a century sound tools of rational choice have been widely available and used, notably PDA (prescriptive decision analysis), involving the quantification of personal judgments of uncertainty and preference. These tools have certainly had some success, and I am confident that structured decision aiding in general has a promising future.

I was a junior member of the Raiffa-Schlaifer team that developed PDA at Harvard Business School in the 1960s. We were convinced these tools would revolutionize how people everywhere go about their business; but so far, it hasn't happened.

Twenty years ago, the US National Academy of Sciences had leading decision scientists study the effectiveness of risk analysis and decision making techniques (Simon, 1988). They reported that the use of PDA and other decision aiding tools was still negligible compared with the great need and potential for them. Since then, the situation has improved, but not dramatically, even for PDA, which is arguably the most promising form of decision aid.

There are certainly encouraging signs. Numerous PDA success stories in various fields have been reported (Corner & Kirkwood 1991; Keefer et al., 2004; Clemen & Kwit, 2001). However, the reports are typically brief and do not really document how a decider's actions were influenced and how beneficial the results were. My experience (which may not be representative) is that the satisfied clients are often staffers who *commission* the aid, not deciders who stand to benefit. More systematic research is needed to establish the facts (and hopefully suggest what distinguishes more from less successful decision aid). In any case, the successes cannot be more than a small drop in a large potential bucket.

*Keynote address to UNESCO Conference on "Creativity and Innovation in Decision Making and Decision Support," London School of Economics and Political Science, June 30, 2006, rexvbrown@aol.com.

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

There are independent indicators that all is not well. General Motors, which had been in the vanguard of PDA supporters, has backed off (Lieberman, 2002). Harvard Business School, the cradle of PDA, no longer makes it an MBA requirement. (However, Howard Raiffa tells me that HBS is planning to re-introduce PDA into the core curriculum, which would do much to restore its professional standing.)

Credible authorities have expressed to me serious skepticism about PDA. They include: James March, Daniel Kahneman and Herbert Simon, noted descriptive decision theorists, two of them Nobel laureates; Jackson Grayson (1960, 1973), a PDA pioneer who later headed the Federal Price Control Board), Stephen Watson (1992), PDA text-book author and later principal of Henley College of Management; and policy advisors to senior Italian, Russian, British and Israel government officials.[1]

It is true that some of our original Harvard team have become highly successful deciders in business and government.[2] Two of my own students got to head a billion-dollar corporation[3] (which bought out my own decision aiding company, which I suppose is a tribute of sorts). They all told me, however, that they make little explicit use of PDA tools, although they find their decision analysis training helps their *informal* decision-making.

2.2 Useful decision aiding

2.2.1 "Decision aiding"

"Decision aiding" is often used to mean *explicit* use of a quantitative decision model to help someone make a better decision, and here is where progress has been most modest. But if we broaden the interpretation of decision aiding to include *any* use of quantitative models, the picture is more encouraging.[4]

For example, training in decision modeling often enhances a decider's *informal* decision making (as with my Harvard business school colleagues). My own advice to executive clients has usually been informal, but honed, I trust, by my decision analysis training. The decision analysis course I now teach (Brown, 2005c) is designed to educate the *intuition* of deciders-to-be, not to have them rely on formal models in their future professional choices.

[1]Respectively, Edward Luttwak, Ivan Yablokov, Herman Bondi, Yezekiel Dror.

[2]Including Ed Zschau, congressman and company CEO; Andrew Kahr, business strategist cited in Nocera (1994) as "one of the great financial visionaries"; Bob Glauber, Assistant Secretary of Treasury.

[3] Bill Stitt, president, and Jim Edwards, chairman of ICF-Kaiser Inc.

[4] I am not concerned here with decision aiding that does *not* involve prescriptive models. They include qualitative techniques, such as lateral thinking and group brainstorming; and decision support systems that do not indicate a specific choice, such as computerized management information systems. What I have to say may not apply to these other types of decision aid.

Another productive use of prescriptive quantitative models is to *justify* or communicate a choice; rather than to make the choice in the first place. Much of my decision consulting business has been of this kind, as for others in my field. For example, regulators regularly use decision analysis models to defend in court controversial rulings against conflicting commercial interests.

In any case, I will be illustrating my argument from my own experience, mainly at my previous research and consulting company, DSC (Decision Science Consortium, Inc.). It deals mainly with using PDA on choices among a few clear-cut (if perplexing) options, rather than with complex decision processes, like oil refining,

2.2.2 Useful decision aiding

The usefulness of decision aiding depends most directly, of course, on how sound the decisions it produces are. What do they promise to contribute, at least in the long run, to human welfare? This depends, of course, on whose interests are affected (such as doctors, patients and tax-payers, in the case of medical decisions.)

However, I am *not* counting as useful aiding clients to make up *other people's* minds in the clients' *own* favor. (They tend to cancel projects as soon as they appear to produce the "wrong" answers. The US Navy approached us to "help" Congress decide whether to buy aircraft carriers or bombers. When I insisted that our findings, whatever they proved to be, should be made public, they lost interest in using us!)

2.2.3 Essential requirements

To be at all useful decision aiding must meet certain essential behavioral and logical requirements. It must:

1. Address the decider's real concerns.
2. Draw on all the knowledge he has.
3. Represent reality accurately.
4. Call for input that people can provide.
5. Produce output that the decider can use.
6. Fit the institutional context.

Decision aiding is useless if *any* of these essentials is lacking, which is often the case.

2.3 Impediments to useful aiding

The main impediments to useful aiding are deficient methodology and its misapplication.

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

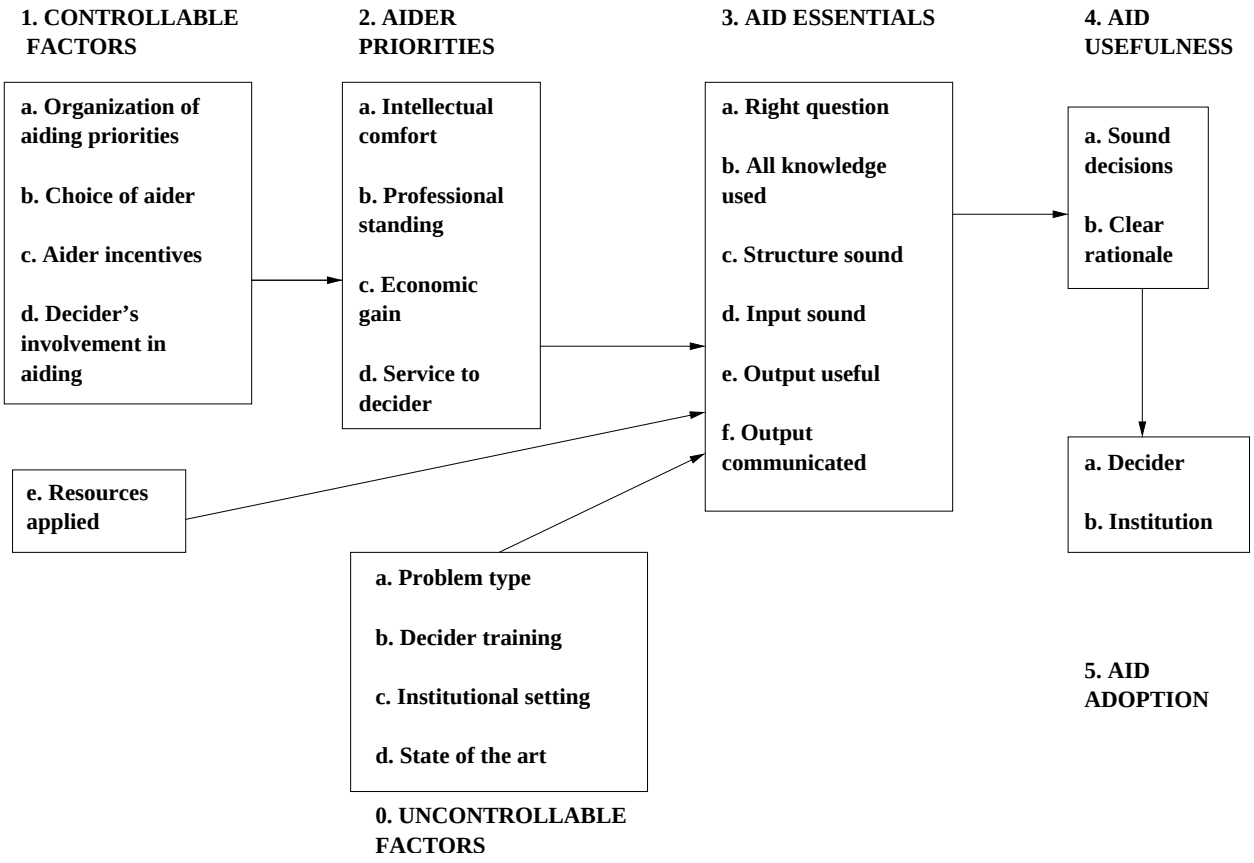

Figure 1: Effect of aider priorities on decision aid usefulness (from Brown, 2005a).

2.3.1 Aider priorities

I have addressed the misapplication impediment in a companion paper (Brown, 2005a). I argued there that aid is often misapplied because decision aiders do not give high priority to being useful. They are under little pressure to do so and therefore to assure that all those essential usefulness requirements are met.

Figure 1 shows the structure of that argument. Whether some decision aid is useful, and therefore adopted (last column on right) is influenced by whether all essential requirements are met (column three). This, in turn, is significantly influenced by the aider's priorities (column two), such as intellectual comfort and professional standing. Aider priorities can be partly controlled (column one), for example by how the aiding is organized and who the aiders are.

2.3.2 Ford depot case

A number of cases in Brown (2005a) illustrate the harm that misaligned aider priorities can do. They include plenty of *recent* failure stories; but I will cite you here an old one which shows with stark clarity, what *can* go wrong. Deciders are not so easily led astray today, because they have learned to be more wary; but the same sort of thing still goes on, in a less egregious form. The case also has a certain piquancy for this audience, because our host, LSE, was involved (though no-one who is still here).

Ford UK suspected it had too many parts depots in South Eastern England, and engaged an LSE operational research group to advise them. The group developed a sophisticated transportation model, which determined an "optimal" number and location of depots. It indicated that, of the seven existing depots, three should be closed. Ford trustingly did so, with disastrous results. The capacity of the four remaining depots proved so inadequate for demand that trucks had to circle the depots endlessly, waiting for space to open up.

It turned out that the analysts had used fatally flawed input (requirement 3d). They had calculated depot capacity as width-times-height-times-breadth, in effect treating it as an empty box to be filled to the top, ignoring unavoidable dead space. They could easily have avoided this gross capacity overestimation by checking with any Ford stock controller. But getting that input right may have been a lower priority than technical satisfaction, and not worth diverting much effort to.

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

3 Research on decision aiding art

Today, however, I want to concentrate on the first impediment to useful decision aiding, inadequate state-of-the art, and to reflect on how decision research could help remove it. Just as aiding decisions has not been the main motivation for "decision aiding," so making decision aid useful has not been the main motivation for decision research.

3.1 Attractive vs. needed research

3.1.1 The record

Following the disappointing Academy report on decision aiding practice that I referred to, DSC got prominent decision scientists and decision aiders together to review the actual and potential impact of decision research on decision aiding (Tolcott & Holt 1988).

The results were disturbing. The participants did report productive descriptive research on how people *do* make decisions and normative research on how they *would* make decisions if they were logical. But they had trouble thinking of recent research that had done much to advance the applied art of decision aiding, with the major exception of influence diagrams. Nor could they cite much research that was addressing problems that decision aiders were *currently* facing.

There have certainly been major innovations in decision aiding technique, but they have tended to come from practitioners. For example, in the 1970s decision aid pioneer Cam Peterson introduced the social dynamic technique of decision conferencing, which is now widely used in business. Academic researchers, however, have often followed through on these innovations, for example, Olson and Olson (2002) with decision conferencing. (Unfortunately hard-pressed practitioner-innovators, such as Peterson, having no academic agenda, rarely publish their work, which would have helped others to build on it. Here is where *aider* motivation works against developing the state-of-the-art.)

3.1.2 Why? Motivation

So, why hasn't decision research been more useful? Richard Feynman once said, "Doing science is like making love. Some good may come of it, but that's not why we do it." The good that may come of decision research is that it improves decisions. The "*Why* we do it" (that is, why researchers do the research that they do) is that it is rewarding professionally and personally. The question is: would more good come of decision research if usefulness *were* the reason we did it? I think so. The other priorities are quite legitimate, but their dominance has created a not particularly useful decision research scene.

(The same imbalance is true of other research fields. The US Department of energy has spent billions (sic) of research dollars on siting a nuclear waste repository. In the course of working on the project, I learned that a major federal research agency was diverting contract money into under-funded research projects more central to their regular scientific mission.)

3.1.3 Gaps

There are serious research gaps in terms of a decision aider's interest (though some may have been filled since I retired from active practice). The gaps are of three types: specialty research; practice-driven research and aid development.

3.2 Specialty research

Specialty research is specific to a discipline, such as statistics or psychology. It is generally convergent, in that: it aims for well-specified and authoritative scientific findings; it usually addresses a *single* aspect of a problem; it seeks universal, rather than topical findings; and it is usually done by university faculty (with their own agendas). Specialty research accounts for most decision research, and certainly produces many useful, even critical, findings, as I will note later. Some research is logical or normative, and some is behavioral or descriptive (which can temper the normative, to produce usefully operational methods).

3.2.1 Logical

Logical specialty research studies what a decider *would* do, if he met certain logical norms, like if he were an "economic man." It includes major work by Savage (1954) on axioms of rationality, Fishburn (1970) on utility theory, Dantzig (1957) on linear programming, and many other models of optimal choice.

Neglected logical topics include:

1. What exactly does decision theory contribute to optimizing choice, beyond testing judgments for consistency?
2. Is there a place in the PDA armory for a construct of *impersonal* probability (Brown 1993)?
3. In everyday life, we progressively develop knowledge about uncertainties in a way that doesn't seem to fit the conventional value-of-information paradigm. Can this common-sense process be productively formalized?
4. How viable is the construct of "ideal" judgment that would result from perfect analysis of a person's available knowledge?

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

3.2.2 Behavioral

Behavioral research *describes* decision processes, including what is wrong with them and why. It includes Tversky and Kahneman (1974) on judgmental biases; March and Simon (1958) on bounded rationality in organizations (1979); and Klein (1997) on naturalistic decision processes.

Neglected behavioral topics include:

1. Systematic review of a sample of past decision-aiding efforts. How good were they? Why were the bad ones bad? What changes might have helped?

2. How can people integrate analytic results into their informal thinking, without disrupting it?

3. We have a good fix on how to make people think smart; but how do we get them to *act* smart?

4. How can *training* in formal analysis educate intuitive and informal decisions?

1. How does the institutional context motivate—or mis-motivate — deciders, decision aiders and decision researchers?

3.3 Practice-driven research

Secondly, there is practice-driven research. This is open-ended exploration of decision-aiding problems and solutions, prompted by lessons learned in the field. The counterpart in medicine is clinical research (contrasted with experimental research). It is divergent in having no predefined end-product, it draws on whatever disciplines the practical need calls for, and often leads to specialty research or aid development (which I will be coming to).

Little practice-driven research gets deliberately planned—or at least funded—mainly, I think, because it is untidy and lacks academic appeal. However, as political analyst George Kennan has said "Tentative solutions to major problems are worth more than definitive solutions to minor problems." It has been argued that practice-driven research will still get done, because researchers will invest in it and get adequate return from fundable follow-up research. However, the researchers in each case are different. Decision aiders are naturals to do practice-driven research (though they may not have the time or qualifications needed.) They produce what I. J. Goode has called *partly-baked* ideas, for specialty researchers to finish baking. (He proposed a Journal of Partly-baked Ideas, where papers were characterized by p as their degree of bakedness.)

DSC was unusually lucky in having an enlightened sponsor at the Office of Naval Research, Marty Tolcott, who was prepared to fund us to do practice-driven research. We interleaved it with our regular decision aiding practice; and it enabled us to prepare a number of successful, more conventional research proposals (in which we often sub-contracted the specialty research parts).

3.4 Aid development research

Thirdly, there is aid development research, which is often prompted by practice-driven research. But, unlike that research, it is convergent, in that it has a clearly defined objective, which facilitates funding. But, unlike specialty research, it addresses the here-and-now rather than the eternal — which *discourages* funding.

3.4.1 Generic

Some aid development is generic and addresses a single aspect of decision methodology. Typically it is carried out by an academic specialist, such as Schachter (1986) on modeling influence diagrams.

Neglected generic questions include:

1. What is the most appropriate form to elicit the utility of a prospect? Should the informant judge utility holistically; or as additive components of utility; or by decomposing such components into factual impact and importance weight, for additive linear MUA (multiattribute utility analysis)?

2. What errors in evaluating options result from common modeling approximations (Brown & Pratt 1996), such as additive linear MUA

3. Empirically, what has the experience of past decision-aiding efforts been? Did they change what the decider did? Did they help, as far as we can tell?

4. How accurately can people make hypothetical factual judgments, both in general and in specific operations, such as the likelihood assessments called for in Bayesian updating?

5. Which decision tools, including non-PDA approaches (such as AHP, traditional OR and behavioral techniques) produce closest to ideal action, when cognitive accessibility, logical soundness and implementation are traded off?

3.4.2 Method-specific

Other aid development is *method-specific*, which focuses on designing a usable tool, such as Henrion's (1991) influence diagram software. Much of it is done by companies who can justify it as a business investment, so funding is less of an issue.

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

It usually takes the form of what design engineers call "build-test-build-test." You use whatever tools you have to solve a problem, see what goes wrong, try to fix the tools, and try them on the next problem. In this spirit, we arranged back-to-back funding from the Nuclear Regulatory Commission (to work on their practical problems), and from the National Science foundation (to develop methodology as needed).

Neglected method-specific questions include:

1. Decision processes commonly consist of *incremental* commitments, but we analyze them as if they were once-and-for-all choices. Is there a practical alternative to cumbersome dynamic programming?

2. How can the reconciliation of plural evaluation models be conveniently computerized (for example, by "jiggling" inputs)?

3. Cam Peterson has a dictum, "Model simple, think complex!" How complex, or structure-intensive, should decision models be, as opposed to judgment-intensive?

4. What is the best balance of decision effort between unaided reasoning based on what you know, getting new information and formal modeling?

3.5 Overall Pattern

My best guess at the proper mix of effort on the three types of decision research, taking into account usefulness and other legitimate criteria, would be to spend about a third on each. More systematic consideration might change this split; but I'd be most surprised if it did not shake the virtual monopoly of specialty research.

An analogy: artistic *evolution* may have turned a simple gothic arch into magnificent Rheims cathedral. But it takes a more pedestrian utilitarian *revolution*, like modular building, to house the masses. In decision research, an evolutionary counterpart would be influence diagrams, where a powerful new idea has been continuously developed over the past 30 years past (I believe) the point of diminishing practical returns, and is still center stage in the PDA world (*Decision Analysis,* 2005). A revolutionary counterpart would be plural evaluation, whose present primitive development (Brown and Lindley 1986) may achieve most of what a greatly refined version could do.

4 Considerations in evaluating usefulness

The research suggestions I am making are based largely on intuitive judgment. Systematic, but still informal, study is needed to check them out and firm them up. It would have to address causal links between research projects and human welfare.

Figure 2 presents a schematic scheme of such causal links. Starting at the bottom, it addresses questions like:

1. What decision tools will a given research project enhance, and how? By improving the tool or how it is applied? ("Direct research impacts" row in Figure 2)

2. How much room for improvement is there in existing decision aiding or in the decision practices it aids? That is, how deficient are tools and practices now? ("Prescription factor" row)

3. As used now, how much will the tools reduce any logical or behavioral deficiencies in "prescription quality" or in "action on prescription"? ("Action factor" row)

4. Will the project only benefit "action quality" or also, say, "cost" (of the decision process) or "institutional values"? ("Benefit type").

5. How do benefits to various classes of decision and population aggregate into total "human welfare"? (Top four rows)

How the various items combine is important. The top levels are usually independent and additive, weighted by importance. At lower levels, however, item contributions may be dependent and non-additive. For example, "prescription quality" and the degree of "action on prescription" may need to be multiplied (rather than added) to get "action quality."

For all its complexity, Figure 2 is by no means complete. It does not, for example, address the usefulness of seeding future research. Nor does it account for who is doing the evaluating. For example, a responsible citizen may consider that a project that improves environmental management world-wide just a little is more useful than research that helps a businessman to prospect for oil a great deal. The American Petroleum Institute may not agree.

4.1 Adapting evaluation to the nature of research options

We only need to consider those items in the causal scheme that are affected by a particular project evaluation.

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

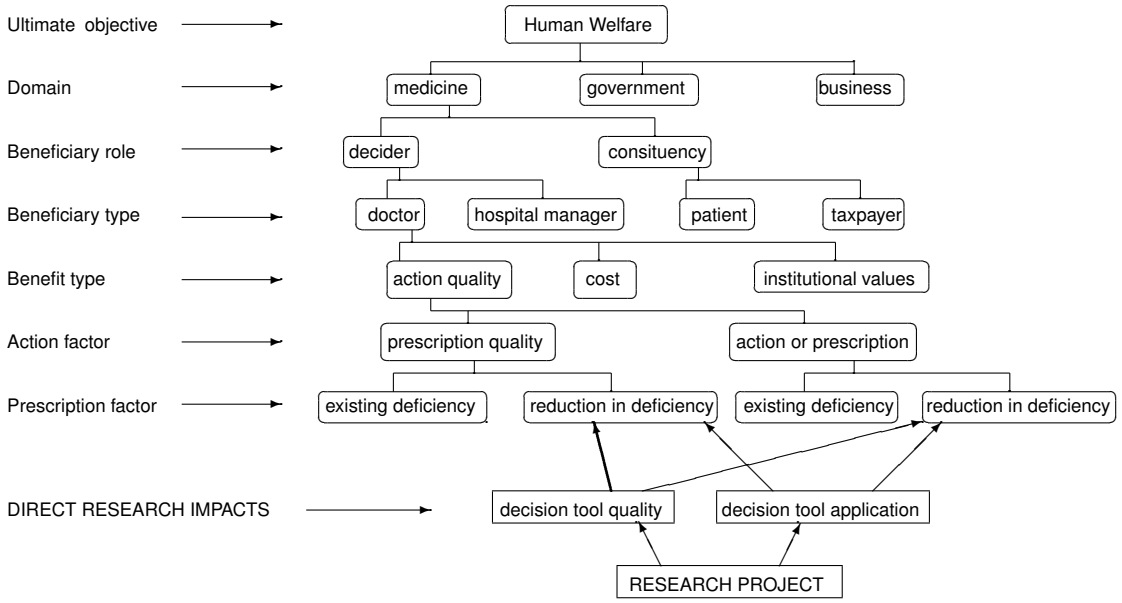

Figure 2: Contributors to research usefulness

4.1.1 Designing a single tool

Suppose the contending projects simply address different aspects of the same aiding tool. The research options may only affect the quality of the choices that this tool prescribes. In that case, we only need to judge which research option improves "prescription quality" most. Attention can thus be limited to the two arrows at bottom left of Figure 2.

4.1.2 Comparing dissimilar projects

However, if research options are more dissimilar than this, more of the causal scheme needs to be considered. Suppose projects address the same decision task, but different aiding tools. (One project may study decision conferencing and the other expert systems, both for medical therapy purposes). Suppose, further, that the tool choice affects not just "prescription quality," but also "action on prescription" and on "institutional values" (e.g., communication). Then all four of the bottom rows of Figure 2 will be affected.

Taking dissimilarity among projects further, suppose they address different domains, different decision tasks and different tools. The choice might be between research on recognition-primed decision for nuclear risk management and research on career planning for the deaf.) Then virtually the whole of the causal scheme would need to be addressed.

5 Quantifying usefulness

5.1 The value of a measure of research usefulness

How is the appropriate refocusing of research effort to be achieved?

5.1.1 Inadequacy of exhortation

Publicizing the above informal reasoning on research usefulness might be all that is required to stimulate useful research practice. In fact, I originally thought that all decision aiders had to do was to tell researchers what research we needed and wait for it to get done. I campaigned rather vigorously for a reformed research agenda, by pitching it to decision science groups[5] around the USA and by publishing articles in psychology and operations research journals (Brown 1989; Brown & Vari 1992). Not much came of it. My issues were not the researchers' issues; and at DSC we were not in a position to do much of the research ourselves. Exhortation is not enough.

5.1.2 Need for motivation

Motivation is therefore needed. The decision research community has had the luxury of indulging priorities other than usefulness, because it could get away with it. I am now convinced that decision researchers, funders and

[5]These included Harvard, Stanford, Duke, Wharton and Carnegie-Mellon.

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

journal editors will only pay real attention to usefulness if they are held accountable for it—or at least get credit for it.

The National Science Foundation does have its proposal referees comment on something like usefulness, under the heading "issue importance." But this criterion is swamped by others, such as technical soundness and originality, and, since the evaluation is qualitative, it does not constrain referees much in selecting proposals.

5.2 Grading research projects on usefulness

I now believe that nothing short of reporting a credible and highly visible *quantitative* measure of research usefulness will move researchers and sponsors to take it seriously. The purpose of the measure would be not so much to improve informal judgment of research usefulness, as to communicate and justify the judgment to others.

5.2.1 Existing precedent

There is some limited precedent for funders giving credit for a quantitative measure of usefulness. NSF's SBIR (Small Business Innovation Research) program does have referees score proposals on usefulness on a five-point scale, under the heading "anticipated technical and economic benefits." This score is added to scores for four more conventional academic criteria. This is fine. But I would like to see that practice extend to *all* decision research procurements.

5.2.2 Credible usefulness measures

The first step in any quantification is to specify the measure. For many research planning situations, such as comparing small-budget proposals, the loose measure SBIR uses may be sufficient. However, more precise measures are called for in high stakes evaluations, especially where usefulness has to be traded off against other criteria. A critical consideration would be whether the user of the evaluation can understand the measure and check it intuitively for plausibility.

A natural metric (like money) may be the most promising usefulness measure. It could be the maximum that the evaluator would consider paying. A funding agency officer might say "The most I could approve awarding for this proposal is $50k. They're asking $100k, so I'm declining it." However, there may be no natural measure that fits the circumstances.

The default measure could be an all-purpose rating scale. The end-points might be zero for present performance and 100 for some ideal. The range of the scale would be the room for improvement in existing aid. The upper end-point could be a project that produces a perfect decision aid (that is, one that makes perfect use of the decider's knowledge), or the greatest contribution that *any* decision aid could make. For example, in a medical context the evaluator might reason: "I project that this research will move the quality of surgical decisions 10% of the way from current practice to some ideal." I am not sure how well such a measure would work, but I will be trying one out presently.

5.2.3 Quantifying the measure

Any measure, however defined, can be evaluated holistically with direct judgment, and that will often be enough. It *could* be derived from a decision analysis model; but I would not give that high priority. The content of the evaluation, and the very fact of quantifying it, is usually more important than how convincingly it is quantified.

6 Real examples

To be more concrete, here is a couple of real research planning choices that I have had to make, with some thoughts on how they might be evaluated.

6.1 Different aspects of one tool

6.1.1 Elicitation vs. logic for Bayesian plural evaluation

My first example involves only the bottom of the causal scheme, and is one of the simpler examples of how a research planning choice might be quantified (if it were worth the trouble). It is a comparison of two method-specific aid development projects. I was preparing a research proposal to develop a Bayesian tool for plural evaluation (that is, making a judgment different ways and reconciling the results). The research design issue was whether to refine the *logic* of an existing model or to improve the *elicitation* of inputs.

6.1.2 Informal Evaluation

I decided in favor of elicitation, on the following informal grounds. Bayesian updating in its current form is almost useless for enhancing intuitive plural evaluation, because people can't provide the likelihood assessments it needs as input. On the other hand, the logic is already quite passable and has only modest room for improvement. Most of the tool deficiency would be cured if elicitation were effective. Since we could make comparable improvements in either aspect for the same cost, elicitation research appears more cost effective.

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

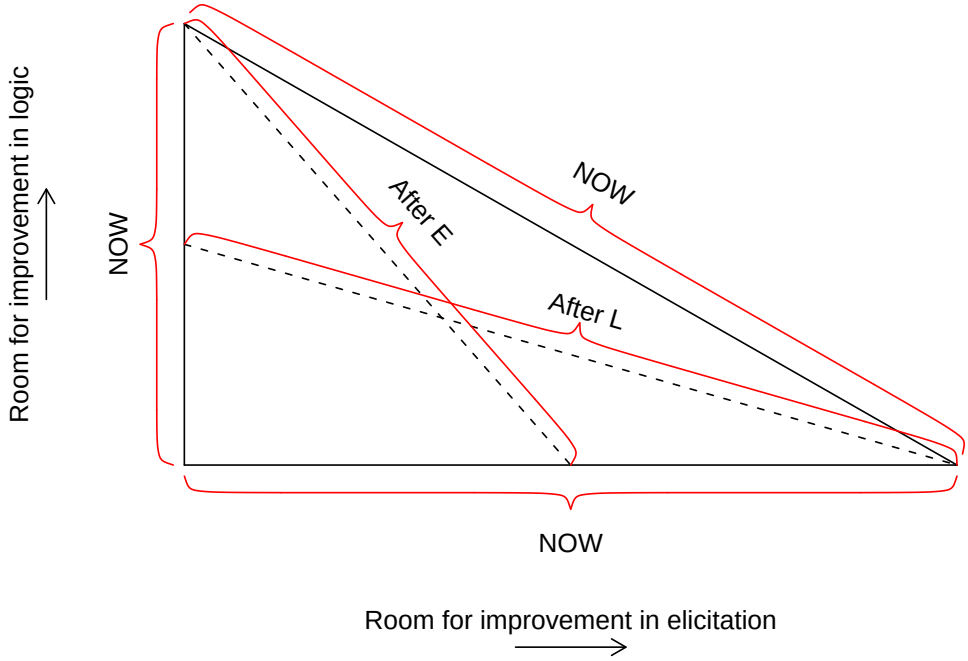

Figure 3: Research on different aspects of a decision tool

My original impulse, however, had been to work on the logic, because my background made me more comfortable with the decision theory involved in the logic issue than with the psychology of elicitation. Moreover, a logic study would give us a better chance getting funded and getting published. In effect, I was swayed by the same distorting priorities that I have been imputing to others. I managed to overcome that impulse.

6.1.3 Quantified evaluation

If we had quantified this reasoning, the measure of usefulness could be "potential improvement in plural evaluation." We could judge directly which project scored higher; or try something more ambitious, like the following.

Imagine an ideal plural evaluation methodology where the modeling and elicitation deal perfectly with what the evaluator knows. Now consider how far short of this ideal the present state-of-the-art falls, i.e. the room for improvement. It seems to me that elicitation and logic relate to that deficiency in a roughly Pythagorean way (rather than, say, multiplicatively). Figure 3 shows that relationship in the context of a right-angle triangle.

The triangle sides are deficiencies in the two aspects. The logic side (on the left) is shorter than the elicitation side (at the bottom), reflecting my view that the logic is less deficient. The hypotenuse gives the resulting total deficiency. If the same effort on either aspect cuts its deficiency by half, the new hypotenuse (dashed line) is shortened by about ten times as much for the elicitation as for the logic project, a great advantage.

My other, private reasons for initially favoring logic were not enough to overcome this advantage for elicitation in usefulness, even if this triangle only approximately models my judgment. So, all in all, the elicitation project is clearly preferred.

6.1.4 Comparing proposals

What would I have gained by this exercise in quantifying usefulness? In this case, probably not very much. It would confirm my informal planning choice and remove any indecision, but it still would make no sense to spend much of the research effort on planning how to spend the rest of it. True, it could also have helped *justify* my choice to the research agency, but still probably not enough to bother.

On the other hand, it could be quite worthwhile to ONR, the funding agency, to grade *all* proposals on usefulness along these lines, to help choose among them. However, the measure of usefulness would now need to be located higher up the causal chain, and take into account more than improving one aiding tool. The measure might go as high as contribution to the quality of *all* military decisions. Furthermore, if ONR wanted to take other criteria into account, the measure of usefulness would need to be explicit enough to permit trade-offs.

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

6.2 Different aiding approaches

6.2.1 Decision analysis vs. Organization design

The next example shows how both the researcher and society could benefit from quantification, in the cause of convincing a research sponsor to support more useful research. The case involved research on alternative approaches to improving military tactical decisions. Navy authorities had noted that in fleet exercises, submarine commanders wait far too long to fire their torpedoes, which puts them at great risk of being fired upon first and being destroyed in a real war.

We were charged with developing a decision tool that would help sub commanders to make more rational firing decisions. Our first analyses confirmed that, indeed, the commanders did wait imprudently long to fire. However, when I talked to commanders, I found that the problem was not with rational choice, but (once again) with motivation. They got credit for pinpointing where the enemy sub was, but they were not penalized for taking unjustified risks (which could get them killed). So it was quite rational for a career-oriented officer to delay firing beyond sound military practice.

6.2.2 Informal evaluation

I urged our Navy client to switch our research assignment from decision analysis to an organizational study of their reward system. We argued that benefits to the navy would go beyond this case and could pave the way for a fruitful new research program. However, our informal argument did not prevail, for bureaucratic reasons: our research grant was part of a larger ONR program on operational military decision aids, and this proposed change was out of scope. So we bowed out of the grant (and luckily found support for the organizational research elsewhere).

6.2.3 Quantified evaluation

It is possible that we would have prevailed over the bureaucratic constraints, if we had made a *quantitative* case on usefulness to our client's Navy superiors. The measure of usefulness might be: reduction in the Navy's loss due to mistimed torpedo firing (adjusted for other criteria, such as seeding new research). The supporting rationale — formal or informal — would address how research might actually change the reward system; and, if it did, what its effect would be on torpedo firing behavior.

6.3 High-stakes risk research

My third example presents by far the strongest case for a quantitative measure of research usefulness, indeed one supported by substantial modeling. The example dealt with an immense research program to aid a critical national choice.

I was a consultant to the US Department of Energy on how to spend literally billions of dollars on whether a proposed nuclear waste site was acceptably safe. I proposed an analytic strategy for allocating this money among various research tasks, and re-allocating it when developments indicated (Brown 2005b). When I implemented the strategy, it indicated major reallocation of the original budget. In particular, in the light of unexpected recent evidence, it recommended more research on gas-borne radioactive release and less on water-borne release, which had dominated the research program so far.

The trouble was that this enormous budget was shared among a few large and entrenched research organizations. They jealously guarded their shares, and none of them had an interest in the gas-borne issue. They wielded enough political influence to block any reallocation. I took my informal argument in vain to an independent Technical Review Board appointed by the US President (and was promptly fired by DOE!). I suspect that a well-modeled quantitative argument presented to the US Office of Management and Budget, the final government arbiter, would have been less easily brushed aside. I might even have gone public with it and pressured Congress to intercede.

Fellow decision-aiders on the project have estimated that the Department of Energy has wasted some *5 billion* dollars on this nuclear waste program, over the years (Keeney, 1987). In this light, I wouldn't be surprised if the difference in usefulness between our proposed research plan and the one adopted amounted to tens of millions of dollars. Thus, a convincing measure of research usefulness might us have saved the American tax-payer a great deal of money. Decision analyst Ron Howard has suggested that 2% of stakes involved in any decision should be devoted to analyzing it. In this case, that would justify spending hundreds of thousands of dollars on comparing the usefulness of research plans—provided the results were acted upon!

7 Conclusions

7.1 Main message

In this talk, I have tried to make the following case: *If grading decision research projects becomes general practice, decision research will be radically transformed, and decision aiding might at last become a major force for better decisions throughout society.*

7.2 Work needed

For this to come about, two things need to happen.

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

7.2.1 Aid development

First, the usefulness methodology must be developed. Adopting the simple existing, if rarely used, procedure of scoring projects judgmentally on an undefined scale would certainly be a significant step forward. It could be tried out in the build-test-build-test mode, on live research planning issues, and refined as needed. One refinement would be to develop meaningful and reviewable measurement scales. Beyond that, evaluating projects indirectly by modeling usefulness (as with the triangle), might also be called for, particularly on high stakes or controversial cases (like the nuclear siting example).

As we know, the best can be the enemy of the good. We don't need to wait for better measures of usefulness before we try to use what we have. Amos Tversky put it to me nicely: "You don't need to finish the foundations before you start working on the roof." Perhaps this audience will be moved to pursue such meta-research. I can't promise it will be free of frustration — just useful!

7.2.2 Lobbying

A critical development needed is institutional. How do you get usefulness evaluation adopted as a general requirement, or at least as standard practice, in research planning? How do you persuade research funders to change their award criteria? Lobbying private sources of funding with reasonable argument may do it. But with government funding agencies I see no alternative to aiders and deciders applying political pressure. I doubt that journal editors can be budged much, but if researchers are adequately funded, perhaps it won't matter.

Actually, there was a major, but abortive, move in this direction a few years ago. The National Science Foundation had recruited a new director from industry, Eric Bloch, whose radical mission it was to make all NSF's research more useful (including our tiny decision research piece.) DSC was charged with studying how NSF should modify its funding procedures, so as to foster more useful research. We started by asking program managers what their funding objectives were. They were resistant, to say the least. The head of physics research told me bluntly "We have no objectives!" I took him to mean, "Leave us alone to do our thing, and don't constrain us with explicit objectives." Bloch did not last long at NSF and his usefulness mission was shelved (along with our own assignment).

It remains to be seen if my more limited present mission, to promote grading decision research on usefulness, will be more successful. If it is, we will have taken a big step towards the golden age of decision aiding we dreamed about 40 years ago. It's worth a try.

Thank you.

References

Brown, R. V. (2005a). The Operation was a Success but the Patient Died: Aider priorities affect decision aid usefulness. *Interfaces, 35*. November-December 2005a.

Brown, R. V. (2005b). Logic and motivation in risk research: a nuclear waste test case. *Risk Analysis, 25*, 125–140.

Brown, R. V. (2005c). *Rational choice and judgment*. New York: Wiley.

Brown, R. V. (1993). Impersonal probability as an ideal assessment based on accessible evidence: a viable and practical construct? *Journal of Risk and Uncertainty, 7*, 215–235.

Brown, R. V. (1989). Toward a prescriptive science and technology of decision aiding. *Annals of OR, Volume on Choice under Uncertainty*, *19*, 1989.

Brown, R. V. (1982). *Prescriptive organization theory in the context of submarine combat systems*. Information and Decision Systems Lab, MIT, December.

Brown, R. V. & Pratt J. W. (1996). Normative validity of graphical aids for designing and using estimation studies. In Zeckhauser et al. *Wise Choices.* New York: Wiley.

Brown, R. V. & Vari, A. Towards an agenda for prescriptive decision research. *Acta Psychologica*, 80, 1992.

Clemen, R. & Kwit, R. (2001). The value of decision analysis at Eastman Kodak Company, 1990-1999. *Interfaces, 31*, 74–92.

Corner, J. L. & Kirkwood, C. W. 1991. Decision analysis applications in the operations research literature, 1970 – 1989. *Operations Research.* 39(2) 206-219

Dantzig G. B. Origins of the Simplex method. In Nash (ed.) *A history of scientific computing.* Reading, Ma. 1990.

Decision Analysis. Special Issue on Influence Diagrams. December, 2005.

Fishburn, P. C. *Utility Theory for Decision Making*. New York: Wiley, 1970.

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

Henrion, M. (1991). Toward efficient probabilistic diagnosis in multiple connected belief networks. In Oliver, R. M, & Smith J. Q. (Eds.) *Influence diagrams, belief nets and decision analysis.* New York: Wiley.

House, P. W. (1988). *Rush to policy: Analytic techniques in public sector decision-making.* New Brunswick, NJ: Transaction Books.

Keefer, D. L., Kirkwood, C. W., & Corner, J. L. (2004). Perspective on decision analysis applications 1990-2000. *Decision Analysis, 1*, 4–22.

https://doi.org/10.1017/S1930297500002382 Published online by Cambridge University Press

Keeney, R. L. (1987). An analysis of the portfolio of sites to characterize for selecting a nuclear repository. *Risk Analysis, 7*, 195–218.

Klein, G. A. *Naturalistic Decision Making* Mahwah, NJ. Lawrence Erlbaum Associates. 1997.

Lieberman, J. 2002. *MarketingDecision Analysis at GM*: *Rise and Fall*. Decision Analysis Affinity Group Conference, Las Vegas, NV.

Majone, G., & Quade, E. S. (Eds.) (1980). *Pitfalls of analysis*. New York: Wiley.

March, J. G., and Simon, H. A. (1958). *Organizations*. New York: Wiley.

Nocera, J. (1994). *A piece of the action.* New York: Simon and Schuster.

Olson G. M. & Olson J. S. Groupware and computer-supported cooperative work. *Human Factors and Ergonomics, 2002*

Shachter, R. D. (1986). Evaluating influence diagrams. *Operations Research, 34*, 871–882.

Simon, H. A. (1979). Rational decision making in business organizations. *American Economic Review, 64,* 493–513.

Simon, H. A. (1988). *Report of National Academy of Sciences Panel on Research needs for decision making*. Reprinted in M. A. Tolcott and V. Holt, (Eds), *Impact and potential of decision research on decision aiding.* American Psychological Association, Washington DC.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Tolcott, M. A., Holt, V. (Eds) 1988. *Impact and potential of decision research on decision aiding.* Washington DC: American Psychological Association.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Watson, S. R. 1992. The presumptions of prescription. *Acta Psychologica, 80*, 7–31.