PART V

HELPING PEOPLE TO MAKE GOOD DECISIONS

Earlier papers in this volume focus mainly on descriptive choice theory (actual behavior) and normative decision theory (ideal behavior consistent with certain axioms). Part V leads off with Rex Brown's discussion of prescriptive decision science whose aim is to help people make good decisions and which draws on both the descriptive and normative theories. Ensuing papers offer further perspectives on the meanings of and interrelationships among descriptive, normative, and prescriptive sides of decision making.

TOWARD A PRESCRIPTIVE SCIENCE AND TECHNOLOGY OF DECISION AIDING*

Rex V. BROWN

Decision Science Consortium, Inc., Suite 300, 1895 Preston White Drive, Reston, Virginia 22091, USA

Abstract

Descriptive decision science says how people *do* make up their minds (e.g. as psychological and organizational theory). Normative decision science says how *ideal* people *would* make up their minds (e.g. as statistical decision theory). Prescriptive decision science (PDS) says how people *should* make up their minds (including a distinctive fusion of the descriptive and normative). PDS supports the development and validation of decision-aiding technology, to make it appropriate for specific circumstances, balancing considerations of feasible input, useful output, logical coherence, and cost of implementation. The author argues for a major redirection of effort toward PDS by the decision science community, and suggests promising directions for its development with illustrations from his company's work.

1. Introduction

1.1. PRESCRIPTIVE, NORMATIVE, AND DESCRIPTIVE DECISION SCIENCE

Prescriptive decision science (PDS) is distinct from both normative and descriptive decision science, but selectively fuses the two, much as engineering fuses physics and mathematics. Its distinctive role is to support decision and inference-aiding technology.

PDS should not be confused with "decision prescription", which refers to recommending – or prescribing – a choice. PDS refers to a higher order of the *process* leading to a decision prescription. "Shoot when you see the whites of their eyes!" is a decision prescription, which may or may not have had the benefit of prescriptive decision science (say, in the form of time-of-fire decision analysis).

Normative decision science, by contrast, characterizes ideal processes and is typified by statistical decision theory and mathematical optimization. In its pure form

* Paper presented at ORSA/TIMS Conference, St. Louis, Missouri, October 1987.

it only needs to meet the internal test of logical coherence. It is context-free and need not be unique (i.e. any given problem may be represented by several normative models).

For example, Bayesian updating simply asserts that *if* a prior distribution and likelihood function are adopted, logical consistency (or at least adherence to certain axioms) requires adoption of a certain posterior distribution. There is no necessary primacy between these three sets of judgments: one could just as easily assert that the likelihood function is implied by the prior and posterior. Whether the directly assessed or the indirectly calculated posterior should be adopted in any particular case (i.e. prescribed) cannot be determined normatively.

Descriptive decision science, on the other hand, characterizes the world as it is, independent of any particular decision-aiding procedure. It is typified by psychology and behavioral organizational theory and by anything else that may be relevant to decision aiding, such as local institutional conditions and the availability of data. A major class of descriptive theory, relevant to prescription, is the psychological work of Tversky and others on the limitations of human decision making.

1.2. PAST NEGLECT OF PDS

In the sixties, decision aiding was dominated by normative developments, typified by the statistical decision theory of Raiffa and others at Harvard. It was widely assumed that a sound normative structure would lead directly to prescriptively useful procedures, of the type developed by Schlaifer at Harvard and Howard at Stanford. However, difficulties in applying these early procedures drew attention to the need to address "people" problems as well as logic problems, and in the seventies, gave rise to the beginnings of a strong tradition of behavioral theory, typified by Edwards in psychology and March in organization theory.

Continuing and expanding attempts to apply the original decision methodology led to a more distinctively prescriptive type of research in the eighties, largely by decision analysis practitioners (with their left hand, so to speak), and by a few academics deeply immersed in practice. However, the scale of prescriptive research has remained at a low level, certainly compared with substantial resources devoted to normative and descriptive research. The main causes appear to be twofold: motivation and funding.

Academics are motivated to seek universal truths, as befits conventional science (whereas prescriptive research is intrinsically topical). Career promotion and recognition tends to go with mono-disciplinary work (whereas PDS is intrinsically interdisciplinary). Furthermore, academic hurdles generally are more readily cleared by convergent research, whose scope can be sharply specified in advance (whereas PDS is usually more productively pursued on a divergent, "follow your nose", basis).

Practitioners, on the other hand, are motivated to do prescriptive research, but typically do not have the economic resources (nor always the research skills) to do so, since their regular clients will only pay for the best technology already available to solve their current problem. Until recently, research funding agencies were set up to support only conventional academic research, i.e. normative and descriptive.

This low level of attention to PDS has contributed to the limited practical impact that, according to some, formal decision-aiding technology has had on decision making in business and government (Grayson [17], Ulvila and Brown [25], Brown [3]).

2. Work needed in PDS

So what is to be done to redress this imbalance? The default situation, and it may not be too bad, is an extrapolation of current trends. For instance, innovative practitioners will continue to tinker with new methods, drawing on available normative and descriptive research, and they will, in turn, prompt new developments in those fields. In addition, one can reasonably expect from academia some spontaneously generated, prescriptively-oriented normative, descriptive, and even explicitly prescriptive research.

However, there is scope for more deliberate PDS research. I offer below what appear to be some promising lines of research.

2.1. AID SPECIFIC PDS RESEARCH

One of the most promising paradigms for the development of specific decision-aiding technologies is the principle, well known in engineering design, of "build-test-build-test". In a sense, this is what practitioners do already, as they try new methods: learn from their experience and then come up with yet newer methods, and so on. Valuable though this is, it suffers from a significant limitation.

The practitioner can only be expected to apply, and therefore to test, what appears to be the best, or at least the safest, methodology available at the time. Otherwise he will not be doing the most cost-effective job for his client. He will not be motivated to experiment with high-risk technologies, even though some of them will prove, often with further development, to be the methodologies of choice in the future. For the same reason, the practitioner is not likely to build or test decision technology more thoroughly than is required to solve a particular problem cost-effectively.

More adventurous and thorough "build-test" requires additional resources and freedom from the tyranny of day-to-day problem solving. Only recently has significant funding become available for this purpose, typified by programs within ONR, DARPA, ARI, and now preeminently by NSF's Decision Risk and Management Science Program.

As an organization which does both consulting and research, we at DSC have found that the most fruitful strategy is to interleave the two. We work on solving an ongoing client's problem as best we can, using available technology. In parallel, we devote any resources dedicated to research to developing and testing innovative techniques for solving the same problem. Any enhanced technology that results may, of course, not be available in time to help *this* client. By pursuing these two tracks in tandem, we are in effect following a build-test-build-test sequence.

The "test" component of the sequence will only be on-site (i.e. applied to the client's problem) if one of three conditions hold. First, the research may be mature enough, and have brought the technology far enough along, that in good conscience we can apply it to the client's problem without sacrificing his interests. Second, the test may be unobtrusive, in the sense that the client is not obliged to do anything differently and, therefore, he does not incur any cost of omission or commission (nor, of course, gain any immediate benefit). For example, several analytic approaches may be applied to a problem "on paper", but only the "tried and true" one is actually used by the client. Or third, the client may be prepared to incur some loss or inefficiency from adopting an experimental technique, as his contribution to advancing the state-of-the-art.

In the absence of (or in addition to) on-site testing, off-site experiments can be conducted. They will have many similarities to conventional experiments in descriptive research, but with a significant difference. The treatment, subject, plot, and yield (to use the analogy of agricultural experimentation) must be highly realistic.

This makes them much more expensive and burdensome than the familiar experiments with college students on contrived tasks. These latter are much favored by behavioral decision researchers in academia, and appropriately so in many cases. If the object is to establish universal truths of human cognition and behavior, there is no reason not to seek the most convenient and cheap instances of them to test.

In the appendix we give an example of some aid-specific research we would like to see done. It is aimed at getting the most out of simple multiattribute utility models, which appear to have the potential for revolutionizing decision-aiding practice.

2.2. GENERIC DECISION-AIDING PRINCIPLES

At a level of generality above specific decision aids, there are a number of generic topics susceptible to research, of which we suggest a few below.

2.2.1. Aiding organizational decision making

The methodology of helping an organization to "make" decisions may differ from that developed for personal decision making. A promising PDS project might be to develop a prescriptive theory of organizational decision making which would play a role comparable to the familiar personalized decision theory. It would cover such issues as: multiple actors, goals and constituencies; hierarchical decision processes; internal versus external organizational actions.

2.2.2. Specifying action alternatives (options)

Among the purely technical impediments to effective decision aiding, the most serious source of failure may be improperly specifying the actions to be aided or to be analyzed (which are not necessarily the same). Issues include:

- Is it better to characterize the options in terms of broad mission or a narrow first step? (e.g. Go west, young man or take bus to airport).
- Do you model the option as a static choice, if the reality involves creeping commitment?
- Where in the hierarchy of possible definitions of choice options do you focus the analysis? (e.g. is the President's problem to be treated as: whether to support a Star Wars bill; or whether to give priority to defense over deficit reduction?).

Unfortunately, it is not clear how one systematically researches such a problem, other than by learning through the evolving experience of analysts trying to do it first one way, then another.

2.2.3. Managing the model-decider interface

This is another of those areas which is critical to the success of a decision-aiding exercise, but where a research procedure cannot readily be identified in advance. Indeed, even the results cannot easily be formulated into guidelines (other than unenforceable guidelines, such as "involve the user"). Perhaps there is no easy alternative to the decision aider learning from his mistakes or to his apprenticing to an experienced practitioner. (I am not optimistic about the prospects for an "expert system" here, though it would be fashionable to look into that.)

One area of decider-analysis interface which is researchable is communication: how to get analytic findings usably into the head of the decider. This has divergent and convergent variants.

The typical case in decision analysis, and management science more generally, is that, when you have completed your analysis, however excellently, the decider cannot – or at least does not – extract from it what he needs. He is faced with either accepting a bottom-line conclusion (which may not capture all the considerations of importance to him), or he is faced with the almost impossible task of figuring out what went into the analysis, so that he knows where and how to take issue with it. We have been doing some work sponsored by the National Science Foundation, Decision and Management Sciences Program, to try and come up with improved communication techniques (Brown and Ulvila [7]). However, we have barely scratched the surface.

2.3. VALIDATION METHODOLOGY

A major type of research specific to PDS is the development of methodology for evaluating the appropriateness of any particular decision-aiding technique or technology. Clearly, for prescriptive purposes, there are multiple attributes to be evaluated, including:

- technical "soundness", i.e. the technique should appropriately incorporate all revelant information and judgment that is available to the subject and that may contribute to the quality of the decision;
- cost, including delay and cognitive burden; and
- acceptability, both institutional and psychological.

The appropriate articulation of, and balance between, these attributes (corresponding, in fact, to importance weights in a multiattribute utility analysis) will be highly situation-specific, in that a technique may fit one situation, but not another.

Note that the normative consideration of logical coherence is only one subattribute under "soundness", which also includes, for example, the quality of inputs it calls for. Thus, approaches with logical flaws are not necessarily irredeemably damned, if they can be shown to improve on unaided intuition – particularly if they have an enthusiastic clientele, who balk at the burden of more "respectable" methodology. The Analytic Hierarchy Process (AHP), which some of us have been inclined to sniff at for its normative failings, may be just such a case.

Of these attributes, technical soundness, and the decision quality it should lead to, are the most difficult to establish and require major methodological development. Not only is it difficult to assess how good a decision's outcome is (since we cannot know with certainty what other decisions would have led to it), it is equally difficult to assess what decision a different decision-aiding methodology would have led to, even if we could measure the quality of either decision.

We believe that the validation of prescriptive technology should be primarily *external*, i.e. tools are tested by confrontation with the outside world. However, *internal* validation, such as through tests of logical coherence and consistency with theory and intuition, has tended to dominate the management science literature. (It is certainly easier, and above all less fraught with nasty surprises, for the proud tool developer!) The development of a methodology for the external validation of prescriptive decision technology offers rich but challenging opportunities for the academic research community. By contrast, aid-specific research, including the application of validation techniques, may fit the practitioner/researcher community better. A significant step in the right direction is Smith et al.'s [23] work on cost-benefit for selecting problem-solving strategies.

2.4. EMPIRICAL META-RESEARCH

Meta-research is research which serves to guide the direction of aid-specific research. Aid evaluation, discussed above, is a clear-cut case in point. Another major category is empirical surveys of existing practice. Important issues include:

- Where, in the world of real decision making, is there the greatest room for improvement and, therefore, the greatest need for an effective decision aid (if one could be produced)?
- What has been the experience with decision aids that have been used?
- What appear to be the obstacles and other factors influencing the success of decision aids?

Such questions need fieldwork to resolve, either extensive surveys or intensive case studies. Although of substantial importance to the development of the field of PDS, neither are particularly high-prestige activities in a university context and are in any case probably best conducted by practitioners.

3. PDS research at DSC

3.1. OUR GENERAL EXPERIENCE

PDS research represents a major part of our work at DSC, probably more than half, if we include the development of decision-aiding technology as well as more basic research. (The balance of our work is comprised of applying existing decision-aiding technology to live problems.) In a sense, PDS research was forced upon us, because when we started in business in the seventies, fresh from academic research and teaching, we made two disturbing discoveries.

The first was that the tools that we and others had fashioned in universities were inadequate (at least as they then stood) to help real world decision making. They needed major research and development, and no one in academia seemed to be doing it. So we were forced to undertake it ourselves, or to specify research projects we could subcontract to our university colleagues, through research contracts primarily from DARPA and ONR (military agencies).

The second discovery was that it takes someone thoroughly immersed in the substantive and institutional setting of a client's problem to *apply* decision-aiding technology usefully. This realization left us with two options: either to specialize in a limited number of problem areas if we wanted to aid decision makers directly, or to develop the technology for others to use, who are already specialists in the problem areas.

By and large, we have opted for the second, tool developer role. However, we have kept our hand in with application projects and, as we grow and mature, we are picking up distinctive mastery of a few application areas. Nevertheless, even in those areas we usually find it more convenient to co-opt problem area specialists into our consulting team, on any particular consulting assignment. In the long run, we think the competitive edge in decision consulting will go to individuals who are masters of both decision technology and problem areas. In the short run, these do not exist and we problem-generalists/technique-specialists get a reasonable share of working on real decisions.

This provides us with excellent testbeds for our prescriptive work. In fact, a common pattern is for us to work reasearch and applied projects side by side, with each stimulating and feeding on the others, along the lines noted earlier.

For example, we recently developed (with an NSF/DRMS grant) some technology for presenting information to deciders (see sect. 2.2 above). In parallel, we tried it out in a project for the Nuclear Regulatory Commission to present information to NRC policy-makers, on the basis of which they would decide whether to require a nuclear plant to install a costly safety feature (Brown [2], Brown and Ulvila [7]). Based on that experience, we have proposed a prescriptive methodology for handling second-order assessment uncertainty (Brown [4]).

3.2. EXAMPLE OF AN EVOLVING PDS RESEARCH PROGRAM: ANALYSIS OF PLURAL RESEARCH (APR)

We will now describe our experience of developing a specific PDS research area which is illustrative of many of the issues noted in section 2. We do not, however, have the space here to get into the technical substance of the prescriptive methodology developed.

In the mid-seventies, our applied work for the Federal Energy Administration and other clients persuaded us that the conventional practice of "singular" decision analysis – i.e. basing a recommendation on a single model or analysis of the problem – often conflicts with common sense. However, the alternative of *plural* analysis lacked a theoretically respectable basis, or much in the way of defensible procedures for designing a plural strategy or merging the results. Conventional decision theory provided a test of *coherence* between alternative single analyses (e.g. between aggregated judgment and a disaggregated model), but not much more. This observation propelled us into an interleaved, largely opportunistic, program of research and application, in the build-test-build-test mold, which has extended over a decade.

3.2.1. The initial normative phase

Build 1. Our starting point was normative and adaptive: to see whether there was any variant of existing stastistical theory which could coherently address the problem. To this end, we joined with a statistical decision theorist, Dennis Lindley (then head of statistics at University College, London) to explore possibilities, and came up with a patchwork of ideas, including an extension of Bayesian updating and preposterior analysis (Brown and Lindley [6]). In order to ensure that any procedures developed would be psychologically sound, we invited a psychologist, Amos Tversky (then at the Hebrew University) to join our project.

Test 1. With support from the Office of Naval Research's Engineering Psychology Program, the three of us illustrated the Bayesian updating approach on a specific (but hypothetical) inference problem – assessing the probability of a binary event, using alternative probability models. It seemed to work, and showed that, at least in this simple case, plural analysis enhanced the quality of the target judgment (Lindley et al. [22]).

Build 2a. The next round of the project was to develop the normative basis, by having Lindley adapt the general Bayesian updating paradigm to a variety of special cases (including plural utility evaluations and plural option selection). This was supported by the ONR Mathematical Sciences Division, which continued to support Lindley and others to develop the normative basis of APR (Freeling [16], Lindley [19–21], Laskey [18]).

3.2.2. Real-world testing phase: A submarine combat aid

Build 2b. We had been working with the Navy for some time, to develop concepts for aids to help ship commanders to make operational decisions in combat (Brown et al. [5]). The time had come to demonstrate whether any of these concepts could be turned into an operational aid that the Navy would actually use (Cohen and Brown [11]). We picked submarine warfare as our testbed and identified target ranging – i.e. estimating distance to a target, preparatory to firing a torpedo – as a decision function for which aid was urgently being sought.

We worked with a Naval Laboratory to develop a computerized "range pooling aid", i.e. an aid for combining multiple estimates of range (Cohen and Brown [11], Cohen [12]). It incorporated a Normal Bayesian updating paradigm, developed by Lindley, which incorporated judgments of accuracy and interdependence about the Navy's contending range estimation techniques.

Test 2a. We tried out this preliminary aid informally on Navy personnel, using input data for a hypothetical engagement that they suggested was plausible. Part of their feedback was that, when the multiple input estimates were widely divergent, the pooled output estimate should show much greater uncertainty than the aid indicated. We realized this was because the Normal model failed to account for the fact that divergent estimates should lead us to downgrade the accuracy ascribed to input estimates.

Build 3. Lindley developed an alternative model based on the *t*-distribution, which corrected the above inappropriate assumptions and produced plausible results (Lindley [21], Ulvila and Brown [24]). Two approaches to validating the aid, and its underlying principles, were attempted.

Test 3a. The first was a statistical outcome-based test. It compared the accuracy of range estimates actually made by a submarine commander during a documented field exercise with the accuracy of estimates which the aid *would have* produced, based on the same available data. This exercise suggested the aid would have yielded significant improvement – of the order of a 20% reduction in average error (Cohen [13]).

The exercise was essentially normative, based on untested (if plausible) ideal behavioral and institutional assumptions about how the aid would be used. (We and the Navy were skeptical that an experienced commander could be improved upon so easily.)

Test 3b. To achieve descriptive realism, the next validation exercise was to replicate, as accurately as a shore-based laboratory would allow, the circumstances of actual application of the aid. The subjects were real submarine commanders and they were led through a real historical combat scenario from a fleet exercise. They were presented with visual and numerical data that they would expect to have, and with realistic timing, in a real engagement.

Removed as this experiment was from the traditional sophomore-in-a-contrived-situation, we still fell substantially short of adequate realism, i.e. simulating a real war, using a fully developed decision aid, under the operative cognitive and motivational circumstances. The results of this "experiment" were largely impressionistic, but provided significant insights into: what an aid of the type we were exploring required of a user; what value it could give to him in return; and what promising directions for improvement would be (Bromage et al. [1]).

Build 4? At this stage, any further development of the range pooling aid was out of our hands. Our role, and that of the Office of Naval Research, our sponsor, is limited to demonstrating the promise of an operational concept. Any steps toward adoption by the Navy, and ultimately installation on submarines, would – properly – be the responsibility of a dedicated Navy organization.

Related efforts. A by-product of the range pooling aid work (which generalized and applied some of its lessons) was a separate build-test-build cycle involving a submarine attack planning aid, and including support for the time-of-fire decision (Cohen [12], Cohen et al. [14]). Our attention has recently been redirected into more basic research (again funded by ONR Psychological Sciences Division). We are now conducting systematic experiments, also involving real Naval subjects in simulated combat environments, to establish more generalized human factors-oriented design-aiding principles. These principles are being incorporated into the design of the attack planning aid (Cohen et al. [15]).

In parallel, we have exploited the familiarity we had obtained with the substance of submarine combat decision making, by working with a major industrial contractor to the Navy, in a strictly applied mode, to help them design a submarine attack planning aid (Ulvila and Brown [24]).

Neither of these efforts directly involved APR, although they flowed organically from the work on it, as has often been the case in our development of prescriptive analytic technology.

3.2.3. Intensive research phase: NSF project on APR

By the early eighties, we were routinely *doing* plural research, having been satisfied by our research and experience that it was generally sound practice. However, we still did not have a prescriptive (i.e. useful) analytic methodology to help us *do it right* – and do it defensibly. Our design of plural research was still largely intuitive.

In 1984, we obtained a 3-year grant from the Decision, Risk and Management Science Program of NSF to develop prescriptive methodology to support the plural research process.

Plural research has two phases, each with two major variants:

- design, or pre-APR, which can be either *static* (i.e. defined all at one time), or *dynamic* (i.e. sequentially developed); and
- use, or post-APR, (i.e. merging plural results), which can take the form of either *reconciling* the approaches (by adjusting their content) or *pooling* the outputs (without changing them individually).

Test Bayesian paradigms. The first step in this phase was to evaluate, from a prescriptive point of view, the then-current state-of-the-art of APR, in all four variants. We found that virtually all the published work – ours and others' – was still largely normative. (A significant exception was pooling plural forecasts (Winkler and Makridakis [27]), where an operational methodology was emerging.) Most approaches were too burdensome to justify using in the course of an applied decision-aiding project, especially in the design phase. It clearly is not worth spending half the available budget to help spend the remaining half!

In particular, our own Pre-APR and reconciliation Post-APR developments had been based on normative Bayesian updating and preposterior paradigms. Our attempts to use them ran into two serious problems deriving from inconsistency with descriptive reality.

First, a major input requirement was a likelihood function. This involved difficult and burdensome second-order probability judgments, which we found too few humans (including statisticians) could supply. More generally, we have found that model-intensive aids, such as this one would be, are extremely difficult and costly to implement validly. Accordingly, we looked for a more intuition-intensive approach which would capture the same notions, but in a more cognitively accessible way.

Build: "feedback adjustment" aid for Post-APR pooling. This led us to develop the notion of "feedback adjustment", for a situation requiring reconciliation Post-APR (Brown and Lindley [9], Brown et al. [8]). Suppose a subject has analyzed a judgmental task (whether of inference, evaluation, or choice), using two or more decision analytic models. If he were coherent, these would produce the same judgment as output, but in general they do not. The Post-APR task is to reconcile the several models, so that their outputs do coincide.

A Bayesian paradigm would have had him assess something like a prior on the components of each model and a likelihood function that reflects the conditional probabilities of his component assessments, given their "true" values. This is what we argue is impracticably difficult in most situations.

Instead, the feedback adjustment procedure has him manipulate all the input assessments, at will, observing the degree of inconsistency in the resulting plural outputs, until those outputs coincide, in which case coherence will have been achieved. If we are lucky, he will properly, but informally, take into account the same higher-order judgments he would have needed for a formal Bayesian solution, for example, about how relatively "firm" his input assessments are and any firmness dependence between them (i.e. corresponding to a joint likelihood function).

Testing pooling and reconciliation aids with controlled experiments. Under the same NSF grant, Detlof von Winterfeldt and his associates at USC have been investigating how subjects pool and reconcile plural estimates *with and without* analytic support (notable feedback adjustment), and with what success. They have adopted the traditional psychological experimental mode, i.e. using students as subjects and giving them hypothetical tasks. They have conducted three waves of experiments, themselves in a build-test-build-test mode (von Winterfeldt [26]). In addition to testing APR principles, these experiments also cast light on implicit plural analysis models in human cognition, which may prompt new or improved formal aids.

Testing feedback adjustment in use. In 1987, we started a 3-year program for the Social Measurement Division of NSF to test the application of this idea in a particular context (industrial market research). Specifically, we are developing software for a market estimation problem (Brown et al. [10]).

We are testing whether researchers and users in the industrial market research community can successfully operate such a feedback adjustment aid and produce acceptable results. (At a later stage, more conventional descriptive research might experimentally test the general properties of the aid in use, such as whether it produces unique reconciliations for a given subject.)

Build a Post-APR design aid. A major impediment to normative Bayesian paradigms being prescriptively appropriate is the difficulty most subjects have in supplying statistical measures of their "confidence" in the plural approaches – in particular, variances and covariances of second-order probability distributions. A DSC member (Anthony Freeling), in the course of an applied plural research project, suggested that the required judgments about the contending plural approaches might be characterized to the subjects as "the amount of information" each contained and be represented graphically by circles, whose relative areas would correspond to variances. "Overlapping information" could be represented by intersecting areas and correspond to covariance.

A prescriptively promising pre-APR design aid would be to compare alternative plural mixes of approaches according to the area of the union of the circles, as elicited.

Test the Post-APR design aid. Whether intersecting areas can adequately provide the basis for such an aid depends on the resolution of both descriptive and normative issues. What subjects' implicit interpretations of the intersecting areas is, is a descriptive issue amenable, with some ingenuity, to experimentation. What formal interpretation *needs* to be put on them, in order for conclusions from the output to be logically valid, is a normative question which mathematical statistician John Pratt, of Harvard, is exploring for us. (Another descriptive issue, less critical but easier to measure, is how good people are at comparing the size of differently shaped areas.)

The general effect of our APR prescriptive research program is that we have some promising and partially developed analytic aids and clearly developed lines for further prescriptive research, consistent with the basic paradigm "build-test-build-test".

4. Conclusions

My general conclusion is that, in order for decision science to have a more beneficial impact in the practical world of affairs, a variety of players need to pay much greater attention to prescriptive decision science viz.:

- Academics need to address normative and descriptive projects shown by meta-research to be most urgently called for and work on distinctively PDS projects such as the development of validation methodology.
- Practitioners need to reserve time (hopefully funded by research agencies) to distill for others what their practice has taught them about PDS.
- Educators need to produce multidisciplinary, application-oriented graduate students (and upgrade their textbooks).

- Funding agencies need to follow the lead of DRMS in promoting PDS (including guidance to reviewers not to apply inappropriate tests of normative and descriptive science).
- Editors need to encourage and accept divergent research (as per I.J. Good's suggestion of a Journal of Partly Baked Ideas) and not apply tests appropriate for convergent research (such as significance at the 0.05 level).

Acknowledgements

This work was supported by the Decision Risk and Management Science Program of the National Science Foundation under Grant ISI-8420732. The author thanks Marvin S. Cohen for reviewing the paper for factual accuracy, and supplying valuable specifics on DSC work referenced.

References

[1] R.C. Bromage, R.V. Brown, J.O. Chinnis, Jr., M.S. Cohen and J.W. Ulvila, Decision aids for submarine command and control, Phase III. Concept implementation, Technical Report 83-2, Decision Science Consortium, Inc., Falls Church, VA (1983).

[2] R.V. Brown, Presenting risk management information to policy makers, Executive summary, Technical Report 85-4, Report to National Science Foundation. Decision Science Consortium, Inc., Falls Church, VA (1985).

[3] R.V. Brown, Decision analytic tools in government, in: *Government Infostructures*, ed. K.B. Levitan (Greenwood Press, Westport, CT, 1987).

[4] R.V. Brown, Handling second-order assessment uncertainty: A prescriptive approach to decision aiding, J. Risk and Uncertainty, in press.

[5] R.V. Brown, C.R. Peterson, W.H. Shawcross and J.W. Ulvila, Decision analysis as an element in an operational decision aiding system (Phase II), Technical Report 75-13, Decisions and Designs, Inc., McLean, VA (1975). (NTIS No. AD A018109).

[6] R.V. Brown and D.V. Lindley, Reconciling incoherent judgments (RIJ) – Toward principles of personal rationality, Technical Report TR 78-8-72, Decisions and Designs, Inc., McLean, VA (1978). (NTIS No. AD A059639).

[7] R.V. Brown and J.W. Ulvila, Communicating information to risk managers: Techniques with examples from nuclear regulation, Vol. I: Main report; Vol. II: Supplementary material, Technical Report 85-4, Report to National Science Foundation. Decision Science Consortium, Inc., Falls Church, VA (1985).

[8] R.V. Brown, K.B. Laskey and G.L. Lilien, Institutional estimation: Measuring quantitative features of populations of institutions (Phase I), Technical Report 85-5, Decision Science Consortium, Falls Church, VA (1985). (NTIS No. PB87-216461/A06).

[9] R.V. Brown and D.V. Lindley, Plural analysis: Multiple approaches to quantitative research, Theory and Decision 20(1986)133.

[10] R.V. Brown, J.W. Ulvila and G.L. Lilien, Personalist appraisal of estimates: A methodology for evaluating industrial market research and other institutional estimation studies, Report No. 17, Institute for the Study of Business Markets (1988).

[11] M.S. Cohen and R.V. Brown, Decision support for attack submarine commanders, Technical Report 80-11, Decision Science Consortium, Inc., Falls Church, VA (1980).
[12] M.S. Cohen, Decision support for attack submarine commanders: Target range pooling and attack planning (U), Technical Report 82-1, Decision Science Consortium, Inc., Falls Church, VA (1982). Confidential.
[13] M.S. Cohen, Validation of a target range decision aid (U), Proc. 49th MORS Symp., Alexandria, VA. Military Operations Research Society (1982). Secret.
[14] M.S. Cohen, R.C. Bromage, J.O. Chinnis, Jr., J.W. Payne and J.W. Ulvila, A personalized and prescriptive attack planning decision aid, Technical Report 82-4, Decision Science Consortium, Inc., Falls Church, VA (1982).
[15] M.S. Cohen, M.A. Tolcott, A.W. Martin, J.M. Leddo and J.R. McIntyre, Cognitive strategies in submarine command decision making, Technical Report 87-13, Decision Science Consortium, Inc., Reston, VA (1988).
[16] A.N.S. Freeling, Alternative frameworks for the reconciliation of probability assessments, Technical Report 80-4, Decision Science Consortium, Inc., Falls Church, VA (1980).
[17] C.J. Grayson, Management science and business practice, Harvard Business Review (July-August, 1973).
[18] K.B. Laskey, Design of plural analysis: Allocation of resources between two or more studies, under revision for Operations Research (1989).
[19] D.V. Lindley, Reconciliation of probability distributions, Oper. Res. 31(1983)866.
[20] D.V. Lindley, Reconciliation of discrete probability distributions, in: *Bayesian Statistics*, Vol. 2, ed. J.M. Bernardo et al. (North-Holland, 1985) pp. 375–390.
[21] D.V. Lindley, Using expert advice on a skew judgmental distribution, Oper. Res. 35(5) 1987.
[22] D.V. Lindley, A. Tversky and R.V. Brown, On the reconciliation of probability assessments, J. Roy. Stat. Soc., Series A142(2)(1979)146.
[23] J.F. Smith, T.R. Mitchell and L.R. Beach, A cost-benefit mechanism for selecting problem-solving strategies: Some extensions and empirical tests, Organizational Behavior and Human Performance 29(1982)370.
[24] J.W. Ulvila and R.V. Brown, Attack planning decision aid for submarine command and control, Technical Report 83-4, Decision Science Consortium, Inc., Falls Church, VA (1983).
[25] J.W. Ulvila and R.V. Brown, Decision analysis comes of age, Harvard Business Review (September–October, 1982) 130.
[26] D. von Winterfeldt, T. Eppel and C. Ford, An experimental investigation of plural analysis: Estimating the market share of compact disc players in 1990, Behavioral Science, in press.
[27] R.L. Winkler and S. Makridakis, The combination of forecasts, J. Roy. Stat. Soc. 146 (1983)150.

Appendix

A PROMISING PDS EXERCISE: FINE TUNING AN MUA TOOL

To give a little concreteness to our discussion, let us address a specific prescriptive issue – how best to measure a subject's (e.g. manager's) utility for a multiattributed consequence. Substantial normative and descriptive research has been done (Fishburn [A1], Tversky [A2]). Recently there has also been important prescriptive research, at a broad level; for example, experimental validation of multiattribute utility measurements (von Winterfeldt and Edwards [A3]).

However, the most pressing prescriptive issue in this area, from our practitioner's point of view, has to do with a narrow implementation issue. Perhaps the most widely and successfully used PDA tool is a simple hierarchical additive model. A popular variant* calls for attributes to be scaled from 0 to 100 and for importance weights which compare the importance of equal intervals on the scales.

We have found the critical prescriptive question to be: How should the scales – and derivatively the weights – be interpreted and communicated? The criterion is how well the aid's output approximates that of an "ideal" analysis, suitably defined. (By this test, more fashionable research issues such as the "utility dependence" of subjects seems typically much less critical. Disregarding such dependence appears to rarely change the preferred choice – but that is certainly a researchable PDS issue.)

The main contending practices are to anchor the ends of the scales either: to well-defined "good" and "bad" scenarios; or to the "option-swing" (i.e. on each attribute the performance of the worst option is zero and the best option is 100). Importance weights are commonly elicited without sharp interpretation. For example, the subject is asked to compare *the* importance of attribute A versus attribute B.

The heart of the problem is that the appropriate weight depends critically on which scale has been used. The normatively appropriate interpretation is clear for either scale, but subjects do not seem to be able to use them (or practitioners do not know how to ask the right questions). In any case, subjects tend to give the same weights, whatever the form of question – at least in the context of top management decision conferences, where time and patience are limited.

If option-swing scoring is used, there is clearly something seriously amiss if the subject assigns a high importance weight to an attribute that he believes is not affected by the options in question. This extreme incoherence would be easy to catch, but we often run into weaker, more insidious examples, where an attribute which is *barely* affected is seriously overweighted, with severe distortion to the analysis.

The prescriptively appropriate procedure clearly depends critically on what the subject "really means" when giving an importance weight. My going-in hypothesis is that, when your typical subject gives relative importance weights, he is thinking about the *general variability* of the attribute. For instance, he is comparing comparable credible spans on the attributes, as they exist in the world he is familiar with. For example, suppose he is evaluating potential dates and all the girls he knows are in a narrow band of intelligence (say because they are all in the same Ph.D. program), but highly variable in beauty. He would presumably give a much higher importance weight to beauty than another subject would who only knows beautiful girls (say, because he works at a model agency).

This interpretation will work fine, using option-swing scoring, for cases where the options span the normal range of the subject's experience; but it can come

*It is typified by the HIVIEW software produced by the Decision Analysis Unit at the London School of Economics, and is used routinely in the Cam Peterson tradition of decision conferences.

seriously unglued otherwise. If one of the contending dates is from a completely different world from the one he is used to, option swings on certain attributes may be greater than his implicit weighting implies. Those attributes will then have less influence than they should, and the choice will be biased against the girl with exceptional qualities (or in favor of the girl with exceptional defects).

A promising PDS exercise would have a psychological experimenter develop pointed experiments which, when reviewed by a skilled decision analyst, would suggest how (if at all) to modify this procedure. For example, if the above cognitive hypothesis about weight interpretation is sustained, the attribute scales might be anchored to familiar ranges, rather than either to reference scenarios or to option swings. The subject might be asked to define a hundred on an attribute scale as a value (not necessarily represented among the options being compared) that is "as high as you normally come across" (and similarly for zero). The exact wording, of course, could be adapted to the setting (e.g. 5 and 95 percentiles for subjects who have had statistics).

The practicing decision analyst will, in parallel with any experiments, be developing his own distinctive PDS contribution to this topic. When he tries an MUA model on a particular subject, he may probe to find out what the subject is thinking of when he provides a weight, and dwell on the feedback if the analysis produces a conclusion which the subject finds intuitively unacceptable. The analyst will, of course, be continually tinkering with the procedure (software, input elicitation, model management, interpretation of output) in the course of his general decision analytic practice.

References

[A1] P.C. Fishburn, Independence in utility theory with whole product sets, Oper. Res. 13(1965)28.

[A2] A. Tversky, Additivity, utility, and subjective probability, J. Mathematical Psychology 4 (1967)175.

[A3] D. von Winterfeldt and W. Edwards, *Decision Analysis and Behavioral Research* (Cambridge University Press, New York, 1986).