CHAPTER 8

Decision-Making

Winston R. Sieck and Gary Klein
Klein Associates, A division of Applied Research Associates, Inc., USA

Decision-making is a pervasive, consequential human activity, from laypeople dealing with their own daily life ~~activities~~, to "tactical" professionals such as physicians, lawyers, firefighters, and military commanders whose responsibilities extend beyond themselves, to presidents and other world leaders who are charged with the fate of us all for generations to come. How do such people make their respective decisions? How should they be making them? And how can we close the gap between what is and what should be? These are the questions that decision researchers have been doggedly pursuing from a variety of perspectives.

The purpose of this chapter is to describe in broad terms two of the fields of decision research that have been devoted to tackling applied problems. Our emphasis is on psychological aspects of applied decision-making, although we recognize that decision research is highly interdisciplinary and we hope that researchers from any number of backgrounds will be able to gain some insights from our review. The chapter describes two perspectives on decision research that rarely intersect: the rational choice perspective, and the naturalistic decision-making perspective. These two perspectives are often contrasted at the ideological level, resulting in discussion that quickly degenerates into polemics. In the current chapter, we attempt to re-engage discussion at the methodological level. Specifically, we compare kernel Decision Analysis (DA) and Cognitive Task Analysis (CTA) methodologies employed in the service of each perspective, respectively. The surprising conclusion is that these methods are intended to achieve distinct goals that are more complementary than competing. In particular, our examination resolves the following misperceptions of these methods:

- CTA sanctifies intuition; DA promotes analysis.
- CTA holds up the expert as having reached the apex of decision-making competence; DA strives for a higher ideal.
- Studies demonstrating weakness in expert judgment imply that DA should be used rather than CTA.

Even more surprising is the conclusion that DA and CTA methods have much to offer toward tackling the big problems of their counterpart. Finally, these potential benefits exist, despite important ideological differences that remain to be resolved.

Handbook of Applied Cognition: Second Edition. Edited by Francis Durso.

Before delving into that discussion, however, we first note that there are three distinct communities of practice involved in decision research, each with its own set of aims and perspectives: Decision Analysis, Naturalistic Decision-making, and Behavioral Decision-making. The Decision Analysis and Naturalistic Decision-making communities are most directly focused on applied decision research and have coherent, deeply rooted approaches for tackling applications. These communities are the primary focus of the current chapter, and thus will be described in more detail below. However, we want to briefly note the community of psychological researchers studying "behavioral decision-making" or "judgment and decision-making." As a whole, this community of researchers embraces perhaps the widest perspective on decision research issues, and hence is less coherent in promulgating a particular approach for applied work. Such researchers are most typically employing experimental psychological methods to investigate basic research issues, and affiliate primarily with cognitive or social psychology departments. Many are increasingly found in other academic schools, such as marketing, medicine, and law, and with slightly more applied foci. The primary forum for behavioral decision researchers is the Society for Judgment and Decision-making (www.sjdm.org). The two primary journals for judgment and decision-making research are *Behavioral Decision-making* and *Organizational Behavior and Human Decision Processes*. There is also a very recent online, open-access journal, *Judgment and Decision-making*.

A second preliminary issue is to define what we mean by applied research in decision-making, given that it is the focus of the current chapter. One approach is to say that applied research is prescriptive in nature. That is, decision researchers have traditionally distinguished descriptive and prescriptive questions to ask about decision-making. The descriptive question concerns how people actually go about making decisions. The prescriptive question, in contrast, asks how people should go about making decisions. Hence, at first blush, basic research could be identified with description, whereas applied research involves prescription. However, this neat distinction is too simplistic. First, many basic psychological decision researchers have prescriptive interests. They sometimes pursue these interests by conducting laboratory studies on college students using rather sparse and simplified decision problems.

As an example, Lichtenstein and Fischhoff (1980) conducted a now classic study to determine whether and to what extent probability judgments could be improved with specialized feedback on their judgment performance. Lichtenstein and Fischhoff were addressing a prescriptive question, namely, the efficacy of a specific approach to improving probability judgment performance. Yet, the study materials and elicited judgments do not particularly reflect those of professionals who are faced with probability judgment tasks, and their participants had no prior experience or current real need to perform in these kinds of situations. As another, recent example, Sieck and Arkes (2005) examined ways of decreasing overconfidence in order to increase undergraduate acceptability and usage of a statistical decision aid in a multiple-cue prediction task. Studies such as these might thus be considered as "basic-prescriptive," and illustrate how basic research on prescriptions can illuminate fundamental decision processes.

Another possibility for identifying research as applied has to do with whether the research is conducted with professionals in their job settings. For example, Dawson *et al.* (1993) had physicians provide estimates on three measures of hemodynamic status, as well as their confidence in those estimates. The physician estimates were compared with values obtained from a catheter to determine accuracy (discrimination and calibration), and neither kind of accuracy was particularly good. In this study, the researchers describe the judgment accuracy of professionals, but without a clear and direct purpose of determining how those professional judgments might be improved. Hence, this research should also not be considered as "applied" by our account.

If neither the prescriptive/descriptive distinction nor the participant population determines whether the research is applied or basic, what is the distinction? In our view, decision research is applied if the project is intended to directly benefit decision-makers who have a genuine need to perform in their respective domain. This emphasis on immediate benefit has a further implication. In basic research, the primary criterion of success is scientific validity. There is no consideration given to how feasible it might be to carry out any recommendations emerging from the findings. Scientific validity is also critical in applied research, but it is insufficient. Applied research has an additional constraint – to consider how the results could be put into practice. In the applied world, usability and acceptability are additional criteria by which success of the project must be measured.

The remainder of this chapter proceeds as follows. First, we provide a brief, general overview of Decision Analysis (DA) methods as employed within a rational choice perspective, and Cognitive Task Analysis (CTA) methods used within the Naturalistic Decision-making perspective. From this overview, we introduce four themes that form the basis for comparison of the alternative research methodologies. Next, we describe the DA and CTA methods in some detail, including how each theme is addressed from within the perspective. Basic research needs to support each method are also described in the context of the final theme on the role of the psychologist. Finally, we compare and contrast the methods in terms of the themes, and describe some of the ways that they can inform each other.

OVERVIEW OF METHODOLOGIES

The DA and CTA methodological processes are illustrated for ready comparison in Figure 8.1. As can be seen, both methods are analytical processes. However, the topics and products of the analyses differ. The topic of DA is the decision problem, whereas the topic of

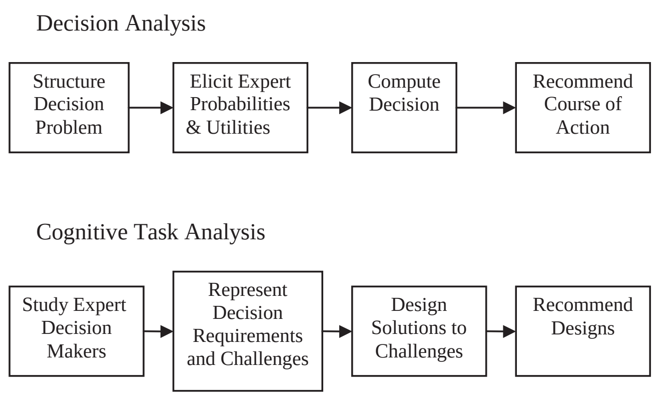

Figure 8.1 Process overviews for decision analysis and cognitive task analysis

E1

CTA is the decision-maker's cognition, including both intuitive and analytical aspects. Also, the product of DA is a recommendation that the decision-maker take a specific course of action. In contrast, CTA produces a set of decision requirements (e.g., the specific difficult decisions for that job), along with designs for interventions such as decision aids or training that support those requirements. As also suggested by Figure 8.1, both DA and CTA researchers are interested in improvement over current practice; neither recommends the status quo or maintains that experts are already achieving their best possible decision performance. At issue are the concepts of quality and improvement themselves. DA relies on formal axioms to define quality and on methods that ensure decisions are consistent with those axioms. For CTA, quality is about how well solutions meet the specifically identified needs of decision-makers, and seeks improvement by bootstrapping expertise. Figure 8.1 also illustrates that experts are incorporated into both kinds of analysis, though the specific roles of the experts differ. Hence, the quality of expert judgment is important for both methods. In CTA, the focus is on judgments actually made by experts on the job, whereas DA elicits quantity judgments that do not necessarily fit the decision-maker's expertise. Given that both methods involve human judgments and other inputs, it is reasonable to ask what roles there are for psychologists. As will be seen, both methods allow for multiple ways in which psychologists could add value to the process, though the psychologist would typically have the lead in CTA. The points of comparison suggested by Figure 8.1 can be summarized in the following themes:

- Topic and product of analysis.
- Quality and achievement.
- Role of the expert decision-maker.
- Role of the psychologist.

Next, we describe DA and CTA in more detail, with an emphasis on these themes.

DECISION ANALYSIS

Decision analysis refers to systematic quantitative approaches for making decisions that
1 are based on a set of formal axioms that guarantee consistent choices (e.g., Clemen 1996). Adherence to the axioms is central to the rational-choice perspective. Decision analysts are most typically associated with operations research engineering, although computer scientists and other computationally oriented researchers are joining the fray. The primary forum for decision analysis research is the Decision Analysis Society of the Institute for Operations Research (OR) and the Management Sciences (www.informs.org/Society/DA). Another group made up primarily of industry practitioners with little interest in academic publishing is the Decision Analysis Affinity Group (DAAG; www.daag.net). The two primary outlets for decision analysis research have been the journals *Interfaces* and *Operations Research*. A new journal, *Decision Analysis*, was launched in 2003. In its preview issue, Keefer *et al*. (2003) provide a thorough review of the field for 1990–2001, including a number of decision analysis trends. A relatively non-technical introduction to
2 decision analysis procedures is provided by Hammond *et al*. (1999). Clemen (1996) and
3 Goodwin and Wright (1998) are two other core texts for the field.

E1

Basic Approach

Decision analyses can become extremely elaborate masses of equations, yet they boil down to three, relatively simple, mathematical rules for normative choice. The first is expected utility maximization, the second is multi-attribute utility, and the third is Bayes' theorem. To describe the rules, suppose that you are faced with a choice between playing a gamble or not. The game is straightforward (at first blush). A six-sided die is rolled. If a 1 or 2 appears, you win $15. If it lands with a 3, 4, or 5, you win $2. If a 6 is rolled, you lose $12. Of course, if you do not play the game, then no money is won or lost; you go home with whatever you had in your pocket. Should you play the game or not?

Before you decide, consider the representation of this decision shown in Figure 8.2. Figure 8.2 is an example of a *decision tree*, and it highlights the various components of your decision problem using typical symbols. The square on the left-hand side denotes a *decision node*, and the two lines running from it represent the available *options*, i.e. to play or not play. The circle lying on the path from the "play" option represents a *chance node*. The lines that stem from the chance node are the *possibilities*. Whereas the lines branching from decision nodes (options) are under your control, those branching from chance nodes (possibilities) are determined by nature. Each possibility has some chance or probability of occurrence. In Figure 8.2, we have listed probabilities for each, assuming that the die is fair. That is, since the chances of obtaining each side of a fair die are 1/6, the chances of a 1 or 2 are 1/3, the chances of a 3, 4, or 5, are 1/2 and the chance of a 6 is 1/6. Finally, the dollar values listed at the ends of each pathway are *outcomes*, and each pathway from decision through to outcome is a *scenario*. That is, each path tells a story about how the situation might play out.

These are the core elements of a decision problem from a decision analysis point of view, and the decision tree is a common means of representing these elements and their interconnections. Now, how would a decision analyst say you should choose? We will start with a simple answer and then add in complications in order to introduce the three rules, and elucidate the basic approach.

The first rule is *expected utility maximization*, in which you should compute the expected utility for each option and then choose the one that is biggest. A simple way to

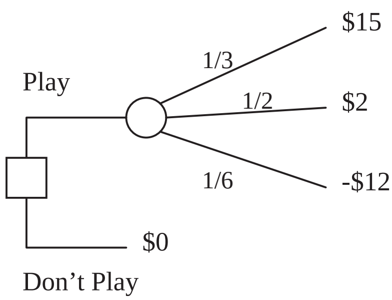

Figure 8.2 Decision tree representation of a gambling problem

do this is to take the dollar value of the gamble as your utility, and compute the expected value (EV) as follows:

$$EV_{play} = 1/3(\$15) + 1/2(\$2) + 1/6(-\$12) = \$5 + \$1 - \$2 = \$4$$

Since the EV of not playing is $0, and $4 > $0, you should play the game, according to this analysis. However, this analysis ignores all the interesting psychological details. First, we simplified your utility or subjective value of each of the outcomes by simply plugging in the objective dollar value. There are at least two ways that these might not line up. First, suppose that you are a teenager, or worse off, a graduate student, and your current total wealth is approximately $14. If, like many people, you are fairly risk-averse, then the possibility of reducing your total wealth to $2 will weigh heavily, whereas the possibilities of ending up with $16 or even $29 might not feel like much improvement over leaving with $14. Hence, your subjective values do not correspond with the objective dollar amounts due to your attitude to risk.

The second complication brings into play the second rule, multi-attribute utility, predicated on the idea that utilities often depend on more than one dimension or attribute. For example, suppose you have a slight moral opposition to gambling. Then your total utility for winning the $15 is determined by:

$$u(\$15, \text{moral failure}) = a \cdot u(\$15) + b \cdot u(\text{moral failure})$$

In this expression, the a and b are weights indicating the importance of each attribute, and the u() indicates that you have to get the money and the feeling of a moral failure onto comparable quantitative scales. That basic mathematical format follows from Multi-Attribute Utility Theory (MAUT).

The final complication involves the probabilities associated with each dollar outcome. We assigned those probabilities based on several assumptions, including that the die was fair. That may have felt reasonable, given the abstract description of the problem. To add some context, suppose that you have encountered this gambling opportunity on a street corner in Brooklyn. The game is being hosted by an unkempt man with an eye patch. It is his die, and he rolls it. He has been on the street corner for a good part of the last three days, though he has not run out of money. As you walked by each day, you saw one play of the game. For the game you saw on the first and third days, he rolled a 6. On the second day, he rolled a 5. It is the fourth day, and you are reflecting on whether or not to stop by the table and play the game. What do you think are the chance of his rolling a 6 (or other number) today? Bayes' rule provides guidance on how you should update your prior beliefs in light of recent evidence to determine a final set of probabilities to include in the expected utility computation. In this case, your priors would be determined in part by your beliefs about chance and probability itself, as well as the situational context. Suppose for simplicity that your subjective probability of a 6, prior to witnessing the three games, is 1/6, but that the context convinces you not to put much stock in that. For the Bayesian analysis, you decide that your prior probability is worth about one real observation. The three rolls that you have witnessed imply that the probability of a 6 is 2/3. Your posterior probability will fall between these two values, specifically it will be: $[1(1/6) + 3(2/3)]/4 \approx \frac{1}{2}$ (Lee 2004). The gamble doesn't look nearly as good at this point.

This example illustrates the basic procedure for conducting a decision analysis, including the components of a decision, the use of a decision tree to graphically represent the problem, and the three fundamental rules for (1) updating beliefs (Bayes' rule); (2) com-

c08.indd 200 9/15/2006 6:20:54 PM

bining outcome attributes into overall utilities (MAUT); and (3) combining beliefs and utilities to determine the overall expected utility of each available option, and then selecting the option with the highest expected utility (EU maximization). However, the simple example does not give much feel for a realistic application of decision analysis. A realistic application is the use for choosing a development strategy for a new drug, Leapogen, at Amgen (Beccue 2001).

In Beccue's study, members of an Amgen leadership team had differing views on the best way to proceed in developing and marketing the drug. They had to resolve their differences and report for a senior management review in two months. The leadership team requested that Amgen's decision sciences group conduct a formal decision analysis to help structure their thinking around the various development options, and to provide a basis for their recommendation to upper management. In response, Beccue formed a cross-sectional team of scientists and managers. They defined eight plausible strategies, and determined the information they would need to complete the analysis. They spent six weeks collecting the data, and built a decision tree with approximately 500,000 scenarios for each strategy. In this case, the optimum marketing focus turned out to be different than any the leadership team had anticipated. Furthermore, the analysis led the team members to unanimous agreement for that under-considered strategy. The team presented the results of the analysis for three of the strategies to senior management, and secured agreement on their recommended strategy from that group.

Topic and Product of Analysis

The topic of decision analysis research is a specific decision problem that requires some resolution. At the end of the process, the decision analysis method delivers the resolution of a decision, or interrelated set of decisions for a particular decision problem of a customer.

Quality and Achievement

The Amgen example illustrates a realistic decision analysis application, as well as providing some insight into how decision analysts think about decision quality. Technically, from the rational choice perspective a good decision is one that satisfies the axioms of consistent choice. This emphasis is seen as an improvement over the idea that a good decision is one that leads to a good outcome. The issue is that the acknowledged element of irreducible chance can produce discrepancies between good decisions and good outcomes. As exemplified in the Amgen example, decision analysts view good decisions as requiring more than satisfaction of the axioms (Edwards *et al*. 1984). Good decisions are typically viewed as those made after comprehensive, analytical deliberation of all options and possibilities. For example, Edwards and Fasolo (2001) described a 19-step process for making a good decision. Their steps include the three normative rules described above, plus supporting processes, such as identifying options and possible outcomes of each option. With eight core strategies and 500,000 scenarios per strategy, Beccue's (2001) analysis certainly exemplifies the state of the art in comprehensive analysis. It also makes clear that extensive time and resources are needed to be thorough in deciding. However, even given that the E1 resource constraints are met, what guarantee is there that such extensive analysis actually leads to better decisions, and in what ways are decisions improved? As with any of the approaches, it is difficult to estimate the value of decision analysis. However, decision analysis would appear to provide at least three concrete benefits: (1) identification of valuable options that were not initially considered; (2) means for teams to negotiate conflicting opinions; and (3) justification and documentation in support of the decision that was made.

As an example, Johnson (2003) and his consulting team used decision analysis to help their clients reach consensus on the next steps to take for a cancer drug. The drug was in use, but there was disagreement over what the subsequent development strategy should be. The decision analysis team identified five key stakeholders in senior management who would approve or veto their recommendation. They interviewed those stakeholders to determine their objectives. In a brainstorming session with a cross-sectional team, they produced five distinct strategies to analyze, but also explored variant, hybrid strategies in the analysis. The decision analysis team discovered that a hybrid strategy was best, with a $50 million NPV improvement over the best of the five original strategies, and $100 million improvement over the status quo (Johnson 2003).

In another example of these benefits, Hammond and Adelman (1976) conducted a decision analysis to resolve a dispute between the Denver Police Department and the ACLU and other concerned citizen groups. The police wanted to change from conventional round-nosed bullets to hollow-point bullets in order to increase the stopping effectiveness, or ability to incapacitate and prevent return-fire. The citizen groups were against the hollow-point bullets because they believed that they caused far more injury. The usual legalistic dialogue was not working well in this case, with each side providing experts to justify their favored bullet and no resolution to the debate in sight. Also, the ballistics experts and politicians did not have clarity on their distinct roles. Hammond and his team were called in at this point to attempt a decision analytic solution. The team assessed subjective values from the mayor, city council, and other elected officials along several dimensions, including stopping effectiveness and injury. They also had the ballistics experts judge many possible bullets in terms of degree of expected stopping effectiveness, injury, etc., based on physical characteristics such as weight and velocity. They then combined the judgments into overall acceptability, using MAUT. Interestingly, the politicians had very diverse subjective values, and ended up settling on use of equal weights for each dimension. From their analysis, Hammond's team found a bullet with better stopping effectiveness and injury properties than those considered by either party initially. The previously unconsidered bullet was accepted by all parties, and implemented by the police (Hammond & Adelman 1976).

These examples illustrate the concrete benefits of decision analysis applications. An example of an application in which conflicting opinions were not resolved involved a study on ecological restoration (Cipollini *et al*. 2005). Ecological restoration is concerned with the management of ecosystems in order to restore them to target levels of ecological integrity. In the current case, members of the management organizations for an Appalachian prairie preserve expressed ongoing disagreement about how to manage all the prairies. The science director incorporated decision analysis, and elicited expert objectives and preferences to develop a prioritization scheme for the prairie management. Cipollini *et al*. (2005) found that, although they had diverse backgrounds, most of the experts bought into the prioritization scheme. However, there was one important individual who did not buy into the scheme (K. A. Cipollini, pers. comm.). This person was the burn crew leader, and so had considerable influence over which management actions would actually be implemented. In the past, this individual had primarily made the decisions based on his own opinion, and was not interested in relinquishing that control to a group process. Further, he believed that decisions about restoration management could not be quantified and modeled. Hence, he put up roadblocks throughout the process, and eventually derailed the implementation.

Pollock and Chen (1986) had similar difficulties implementing decision analysis in China. Specifically, they met with resistance against the analytical approaches, and the participants had considerable difficulty in assessing probabilities and subjective values. These cases highlight that, although decision analysis may aid in resolving conflict, it does not constitute a panacea for negotiating differences, since, for example, parties may well disagree on whether the general approach is acceptable. From an applied research standpoint, it is extremely risky to dismiss such findings or attribute them to the imprudence of the decision-maker. In the applied world, usability and acceptability of the analysis and recommendations are critical, in addition to scientific validity.

Role of the Expert Decision-maker

The subject matter expert's role in the group decision process is to provide particular judgments to support the decision tree (or related) representation, when called upon. In some cases the expert may aid in the initial problem structuring, but this is not a hard requirement. The primary information needed from expert judgment to support decision analysis is a set of probability judgment inputs. For example, the subject matter expert used to populate Hailfinder made 3700 probability judgments (Edwards 1998).

In addition to supplying judgment inputs, the expert typically has some role in reviewing the output of the analysis, and might have some or complete say in whether the recommended option will be accepted.

Role of the Psychologist

There are roles for psychologists to participate directly in the applied decision analysis research, as well as in conducting more basic studies that support decision analysis. In the application studies per se the formal representations and analyses dominate, so that psychologists tend to be placed in supporting roles for the OR engineers. For example, Edwards described the need for skilled quantity judgment elicitors, and elicitation facilitation and training as a primary way for psychologists to contribute (Edwards 1998). Probably a more important role for psychologists is to facilitate the initial framing of the problem, in part to ensure that important aspects from the key players' mental representations are not lost in the rush to formalization. Finally, there is a role for psychologists with knowledge of social/organizational decision-making who can flag and cope with the myriad interpersonal issues that surface throughout the process (Peterson *et al*. 2003).

E1

These requirements for psychologists "in the moment" during conduct of the decision analysis study naturally correspond to the needs for psychologists to carry out supporting basic research. The three points of input for basic research psychologists are described in turn.

Elicitation: Assessing Subjective Probabilities and Values

An obvious role for psychologists conducting basic research to support decision analysis is to research and develop elicitation methods for subjective values and probabilities. Indeed, much basic research in judgment and decision-making has been devoted to discovering new ways to conceptualize and model subjective value and subjective probability. For example, Mellers and her colleagues have developed new models of subjective value that better account for emotional responses (Mellers *et al*. 1999; Mellers 2000). Birnbaum (2004, in press) has conducted a number of tests of several advanced theories of utility. The basic paradigm for this research is to construct mathematical models that accurately predict people's decisions in very transparent situations, and then test their goodness-of-fit as compared with other models. For a decision analysis application, the point would be to use the most psychologically realistic models of subjective value in their analyses of large, "real" decision problems. This approach is predicated on two tacit assumptions: (1) people decide according to their "true" preferences for small, transparent decisions, but not for large decisions that are more opaque; and (2) a highly detailed, accurate model for the transparent decisions is what should be incorporated for use in decision analysis on the opaque decisions. The large decisions are beyond unaided human capacity to jointly consider all of the elements, so the person is "represented" in the equation by a highly realistic mathematical proxy.

The decision analysis examples described above suggest a very different paradigm from the one just described for subjective value elicitation and model development. First, note that the value models were attempting to capture a group's values, not those of any one individual. Also, the people appeared to be more concerned with ensuring that their "position" was represented in the model, and that the various positions are represented in a simple, understandable, and fair way. Hence, a basic line of psychological research on value assessment that would seem to better serve decision analysis would put a group of people in a scenario where they have distinct preferences, and then investigate the group process by which they come to agree on a model of their collective preferences. Note that the criterion here is acceptability, or willingness to use the value model, rather than goodness-of-fit of the model to simple gambles. Most current psychological research appears
4 to be moving in the opposite direction (but see Edwards & ~~Barron~~ 1994).

Subjective probability is another area that has received considerable attention by psychologists, but in a way that seems disconnected from decision analysis. Edwards and others have noted that probability research in the lab~~.~~ is quite different from the way elicitation is done in the field (Edwards 1998). Rather than studying probability judgment using general knowledge questions, for example, researchers ought to focus on future events (Wright & Wisudha 1982). It should also be useful to study judgments of conditional probabilities clustered within an overall scenario to match those required
E1 in decision analyses, rather than investigating judgments of independent items as typified in the lab.

Understanding the Situation

A bigger and more interesting contribution that is needed from basic psychological researchers is in coming to grips with a step that precedes application of the three rules in the decision analysis process, "understanding the situation" (Edwards & Fasolo 2001). Research into understanding falls squarely within the domain of the cognitive psychologists, and presents an ideal phase in which they can make a difference (Weick 1995; Klein *et al*. in press). Understanding the situation, or initial problem structuring, has been held as a crucial and poorly understood component of the process (von Winterfeldt & Edwards 1986). Graphical representations, such as knowledge maps, have been developed and incorporated by decision analysts as an approach to improve problem structuring (Howard 1989). However, situational understanding remains severely underdeveloped in decision analysis, though psychologists have conducted some interesting initial studies (Benson *et al*. 1995; Browne *et al*. 1997).

Organizational Issues

In practice, DA tends to be done with multiple stakeholders. It follows that there is a need to coordinate between analysts, managers, and executives. Decision analysts have developed some of their own processes for managing interactions between stakeholders and decision analysts, and for dealing with multiple decision-makers (Kusnic & Owen 1992; Bodily & Allen 1999). However, as in the case of situational understanding, there is a clear role for psychologists in the development, vetting, validation, and implementation of such processes.

COGNITIVE TASK ANALYSIS

The Naturalistic Decision-making (NDM) perspective focuses on improving decision-making by first investigating how highly experienced individuals and groups make decisions in field settings. NDM researchers investigate decision-making in settings that are marked by ill-defined goals, shifting conditions, high levels of uncertainty, high stakes, non-defined or ill-defined cues, contextual complexity, and organizational constraints. We need to examine decision-making under these kinds of conditions in order to improve the track record of applications. Therefore, NDM researchers have needed to develop methods for cognitive field research, particularly observations and methods for conducting interviews.

Cognitive Task Analysis (CTA) refers to a set of systematic, qualitative methods for uncovering information about the knowledge, thought processes, and goal structures underlying observable performance (Cooke 1994; Militello & Hutton 1998; Schraagen *et al*. 2000; Crandall *et al*. 2006). NDM researchers and CTA practitioners tend to be associated with applied cognitive and human factors psychology. The primary forum for NDM research is the Cognitive Engineering and Decision-making technical group of the Human Factors and Ergonomics Society (cedm.hfes.org), and the bi-annual Naturalistic Decision-making conferences (e.g., www.ndm7.com). The primary outlet for NDM research has been the series of edited volumes resulting from the NDM conferences E1

(Zsambok & Klein 1997). However, the new *Journal of Cognitive Engineering & Decision-making* should provide a good, peer-reviewed forum for NDM and other cognitive field research studies. A non-technical introduction to NDM is provided by Klein (1998). Hutchins (1995) gives another good introduction to the same general concept.

Basic Approach

Cognitive Task Analysis provides a set of knowledge elicitation, analysis, and representation tools for identifying and analyzing general domain knowledge, expert knowledge, and task-specific skill sets for critical decisions made in a subject-matter expert's (SMEs) complex, naturalistic environment. These analyses lead to a breakdown of the tasks and functions into critical decision requirements: the critical, challenging, and frequent decisions that are made, and the information, knowledge, and strategies that go into making those decisions. Those decision requirements are then used to develop decision-skills training, inform the design of decision-support systems, or restructure teams and organizations. Next, we lay out the basic approach, and provide a couple of illustrating examples from actual projects.

There are several specific CTA interview methods, perhaps the most central of which is the Critical Decision Method (CDM) (Klein *et al*. 1989; Hoffman *et al*. 1998; Schraagen *et al*. 2000). CDM was originally developed based on an earlier technique for uncovering critical incidents (Flanagan 1954). A CDM session is organized around an initial account of a specific incident that comes from the participant's direct experience. The specific episode carries context with it and reveals how particular aspects and events in the environment impel the decision-maker to action.

The CDM interview requires an initial step, that of guiding the participant to recall and recount a relevant incident, depending on the focus of the study. The interviewer then conducts three additional information-gathering sweeps through the incident. These are: Timeline Verification and Decision Point Identification; Progressive Deepening; and What if Queries.

First Sweep: Incident Identification and Selection

In accord with the goals of the project, interviewers will have decided on an opening query, such as "Can you think of a time when you were fighting a fire and your skills were really challenged?" The idea is to help the SME identify cases that are non-routine, especially challenging, or difficult.

Once the participant identifies a relevant incident, he or she is asked to recount the episode in its entirety. The interviewer acts as an active listener, asking few questions, and allowing the participant to structure the incident account him or herself. By requesting personal accounts of a specific event and organizing the interview around that account, potential memory biases are minimized (Berntsen & Thomsen 2005).

Once the expert has completed his or her initial recounting of the incident, the interviewer retells the story. The participant is asked to attend to the details and sequence and to correct any errors or gaps in the interviewer's record of the incident. This critical step allows interviewers and participants to arrive at a shared view of the incident.

E1

Second Sweep: Timeline Verification and Decision Point Identification

In this phase of the interview, the expert goes back over the incident account, seeking to structure and organize the account into ordered segments. The expert is asked for the approximate time of key events and turning points within the incident. The elicitor's goal is to capture the salient events within the incident, ordered by time and expressed in terms of the points where important input information was received or acquired, points where decisions were made, and points where actions were taken. These decision points represent critical junctures within the event. At the conclusion of the second sweep through the incident account, the elicitor has produced a verified, refined documentation of events.

Third Sweep: Progressive Deepening and the Story Behind the Story

During the third sweep, the CDM interviewer leads the participant back over each segment of the incident account identified in the second sweep employing probes designed to focus attention on particular aspects of the incident, and soliciting information about them. The data collected in the third sweep may include presence or absence of salient cues and the nature of those cues, assessment of the situation and the basis of that assessment, expectations about how the situation might evolve, goals considered, and options evaluated and chosen.

Fourth Sweep: "What if?" Expert Novice Differences, Decision Errors, and More

The final sweep through the incident provides an opportunity for interviewers to shift perspective, moving away from the participant's actual, lived experience of the event to a more external view. During this phase, interviewers often use a "What if?" strategy. They pose various changes to the incident account and ask the participant to speculate on how things would have played out differently. An example query might be: "What if you had taken action Y, instead of action X?" Answers to such questions can provide important insights into domain-specific expertise. Or one might go back over each decision point and ask the expert to identify potential errors and explain how and why those errors might occur in order to better understand the vulnerabilities and critical junctures within the incident.

After a number of such incidents have been collected (*n* is typically between 8 and 40, depending on the size of the project), qualitative analyses on the transcripts are conducted. Analysis of these data reduces and combines the individual incident accounts into the form of essential "decision requirements." The decision requirements encompass the "whats" and "hows" of decision-making. What are the critical decisions? What pieces of information are critical to making those decisions? Decision requirements summarize the key difficult, challenging, and critical decisions within a specific task domain. These descriptions are supported by a host of relevant information about why the decision or cognitive activity is challenging or difficult, what strategies are used to make the decision, what information or patterns of information support the decision-making, what critical background knowledge is required to support the decision, what common errors or difficulties are incoming to a decision, and so forth. The requirements then provide the basis for organization or technology design concepts, process improvements, and the development of training exercises for decision-making (Klein *et al*. 1997; Hollnagel 2002).

5 One example of a CDM interview is a project conducted by Klein ~~*et al*.~~ (1985) to study how fire ground commanders make life-and-death decisions under extreme time pressure. This was one of the first studies using critical incidents as a basis for interviews, and much of the formal structure of the CDM was developed to facilitate the research. The research team conducted CDM interviews. Each interview yielded one or two critical incidents, and these incidents contained approximately five decision points. The final tally in this study was 156 decision points which were probed. The researchers categorized the way that the fire ground commanders handled each decision point. In 127 instances, the commanders recognized a promising option and adopted it without systematically comparing it to alternatives. In 11 instances, the commanders had to devise a novel course of action because none of the usual methods would work. In 18 instances, the commander recognized two or more promising options and compared them to find the best by imagining what would happen if the course of action was carried out.

Another CDM study was conducted with weather forecasters (Pliske *et al*. 2004). A total of 22 Air Force weather forecasters were interviewed to understand how they generate their predictions. Pliske *et al*. found a large variation in strategies, and hypothesized that the experience level and engagement of the forecaster determined the strategy used. Some forecasters applied formulaic methods to compute their predictions. Others used complex rules. Note that extensive quantitative analysis is among the strategies in the repertoire of these experts, as it fits naturally within the demands and constraints of their jobs. Even so, the expertise of the most skillful forecasters went beyond rendering their predictions. For example, they could identify the "problem of the day" that needed to be closely monitored. As important as this problem detection process is, it has been virtually unstudied. They could also sense when the standard computer projections were not trustworthy. Specifically, they could use scatter in the data as an important indicator, and had a sense of what the necessary level of data-smoothing was in particular situations. They would not use programs that over-smoothed the data, and did not permit the user to adjust the level of smoothing. Without understanding in depth the competence and situations of these experts, the applied decision researcher is at risk of attributing the expert's rejection of such an analysis tool to some inherent character flaw on the expert's part.

6 Hutton ~~*et al*.~~ (2003) described research employing CDM and related CTA methods to understand the cognitive challenges faced by experienced air campaign planners. These challenges included being faced with large amounts of dispersed data, managing conflicting goals, and dealing with unclear criteria for the development and evaluation of plans. Taking the details of these challenges into account, the research team designed and developed a prototype decision support tool, the Bed-Down Critic, for use by the planners in the development and evaluation of their plans. The planner could drag and drop assets onto potential sites (e.g., aircraft onto airfields) and receive immediate feedback on the appropriateness of the allocation, thus alleviating the need to attend to myriad lower-level details and remain focused on the higher-level objectives.

These examples illustrated how a CTA method such as the CDM can be useful for obtaining insights about judgment and decision strategies in complex field conditions, and how those insights can be translated into support systems that are tailored to meet the needs of decision-makers.

E1

Topic and Product of Analysis

The topic of a CTA project is the decision-making process of a set of experts in the context of their work. At the end of the process, CTA delivers insights into the nature of the decision problems from the experts' point of view. More concretely, these insights are captured in the form of decision requirements that guide the design of solutions (e.g., decision aids, training, organizational processes) that fit with the realities of the end-users' situations. There are few a priori constraints on the nature of the recommended solutions; they are truly tailored to fit the decision-makers' requirements.

Quality and Achievement

NDM researchers typically study experts to describe good decision-making processes and define quality decision-making (Klein 1989; Hutton & Klein 1999). Hence, the primary criterion for good decision-making is to act like the experts: to notice the cues and patterns, to make the adaptations, to engage in anticipatory thinking, and so forth. A good decision involves doing these sorts of things in ways that novices would not. If we use expertise as a gold standard, a good decision means complying with the application of expertise. This idea of describing expertise in order to set a gold standard for a domain clearly requires further discussion. For one thing, it brings us back to the description (how do people make decisions) vs. prescription (how should people make decisions) distinction. With that distinction in mind, an immediate question that surfaces is how can we possibly move directly from describing decision-making to prescribing it?

To address this question, consider the analogy to bootstrapping in judgments with specified criterion and cues (Dawes & Corigan 1974). In bootstrapping, the decision researcher builds a descriptive model of judges by capturing their policies in equation form, and then replaces the judge with the equation. The bootstrapped equation is often found to perform better than the original judge. CTA moves from description to prescription in a similar manner, though the endpoint is an aid for the judge, rather than replacement.

In general, CTA studies extract and model the expertise of a group of performers, not individual experts. That pooled information from the group of experts is then used to determine the critical judgments and decisions that the experts actually make on the job, what the challenges are in making those judgments and decisions, and what cues, factors, and strategies they use to make those judgments. Strategies range from intuitive rules of thumb to extensive analytical processes, depending on the constraints of the job. Also, NDM researchers are not drawing random samples in order to describe the characteristics of a general population of performers in a domain. They are attempting to draw samples from the top performers, and to construct descriptive models of the decision-making of that expert group. The resulting descriptive model that culls lessons learned from the set of experts is then used as the gold standard, or current best available practice.

Establishing the cognitive performance of the experts as a gold standard provides several advantages. First, since the description of skilled performance combines cues, strategies, etc. of a group, there are elements of the description that any individual expert might pick up. By clarifying the kinds of cues and relationships that experts are attending to, we can see how to formulate a program for skill development in journeymen and novices. Second, because the model is stable it suggests reliable achievement of the E1 strategies. Hence, individual experts can themselves pursue more reliable application of the described process. This latter point suggests that we might be better off considering the judgment and decision strategy of experts as a baseline reference point than a gold standard. We do not want to place a ceiling or end state on their processes. Rather, we want to take their strategies into account when considering ways to improve processes, in part to ensure that our recommendations are usable and acceptable.

What kinds of impact can a CTA study have? Staszewski (2004) conducted a study of expert minesweepers. He used CTA methods to identify the cognitive strategies they use in detecting small mines with low metal content. For example, he found that experts altered how they swept detectors over the ground, slowing the rate of their sweeps and using the detectors to find the edges and centers of the mines. Their judgment process is more complicated than mapping a set of static cues onto a criterion. Even the newest technology only resulted in a 10–20 per cent success rate in detecting mines. Staszewski used his findings to design a training program to help Army engineers use the strategies of the experts. The results were dramatic. Detection rates rose from 10–20 per cent to 90 per cent and higher. Even an abbreviated training program administered for only an hour resulted in success rates higher than 80 per cent.

Role of the Expert Decision-maker

The subject matter expert (SME) and decision-maker is the focal point of analysis. The point of the research is to foster a collaboration between researcher and SME, as an active participant in the design process. Taking the experts seriously means realizing that they develop their own tools themselves, and recognize what new solutions will help them. In a CTA study, experts also have the roles of sharing what they have learned, articulating what they notice, admitting mistakes where they think they have made them, and speculating about other approaches or reactions.

Role of the Psychologist

As in Decision Analysis, there are direct roles for psychologists in research using CTA to improve decision-making, as well as in basic research that supports the approach. In direct applied CTA research, the psychologist is the driver of the CTA process and design/development of supporting training or technologies. Technology experts, such as OR experts, play a supporting role, rather than a lead role. Several applied and basic research roles for psychologists are described next.

Knowledge Elicitation

Psychologists working within the NDM framework perform cognitive task analyses to identify the environmental factors and pressures that affect decision strategies and quality. The knowledge elicitation that CTA researchers conduct facilitates the initial framing of the problem and ensures that important aspects from the key players' mental representa-

E1

tions are not lost in the rush to develop technological or training solutions. Indeed, CTA methods deliver the kind of broader problem structuring insights that decision analysts have described a need for.

There are also basic supporting research roles. Although CTA methods have been carefully thought out and used in the field for decades, rigorous development of the approach is lacking. Hence, considerable further research is needed to study, develop, and validate CTA and novel methods, especially knowledge elicitation methods, to improve their quality and rigor. For example, Nisbett and Wilson (1977) have criticized the use of verbal reports in decision research by arguing that experts are conscious of the products on their mental processes, but not of the decision processes themselves. CTA researchers are well aware of the limitations of incident recall. Klein (1998) cites two cases where decision-makers believed that they made decisions because of ESP. Not believing that answer, the CTA interviewers emphasized probes about what the participants recalled noticing, rather than asking them to explain their decision strategy. And that procedure uncovered a much more plausible basis for the decisions. Hence, while the Nisbett and Wilson (1977) study cannot be taken as a definitive condemnation of CTA because the specific probes differ, further basic research is clearly needed to determine the real validity constraints on incident-based interview approaches, as well as positive suggestions for methodological improvement. Research into methods for interviewing eyewitnesses is likely to prove a reasonable starting point for such investigations (Geiselman & Fisher 1997).

Expertise

Another clear role for psychologists is to discover the basis for expertise. In basic, supporting research, psychologists are needed to help resolve the process–performance paradox. A first step might be to ensure that the precise judgment selected and task details actually match the critical judgments and details of how they are made on the job. A CTA combined with a modified experimental judgment paradigm might be useful to perform this kind of study. The CTA would be used to identify critical judgments actually made by experts, along with their processing strategies for deriving relevant cues from the environment (e.g., as in Staszewski's minesweeping study). An ecologically valid judgment paradigm that stresses those processes could then be constructed and used for investigation into expert performance. Another related area of importance is the development of assessment approaches for clearly identifying/vetting experts (Shanteau *et al*. 2003).

Solutions Design and Development

Psychologists also have a key role in applying the discoveries made in CTA studies in order to improve performance, whether through training, or information technologies, or incentives, or organizational redesign, or personnel selection. In order to accomplish this application, psychologists have to collaborate with developers of information technology, working to get them to think about the point of view of the expert decision-maker (Hollnagel 2002).

E1

DISCUSSION

Psychologists interested in performing applied decision research might do well to consider how they want to contribute to the broad community interested in improving decision-making. Psychologists conducting basic research might reflect on how they arrived at studying the particular set of issues and paradigms that they are pursuing. What are the originating and extant implications for their work to applied decision research?

To assist psychologists in defining their desired role, in this chapter we have compared and contrasted Decision Analysis (rooted in a rational-choice perspective) with Cognitive Task Analysis (stemming from the Naturalistic Decision-making perspective). Furthermore, we have attempted to identify needs for psychologist input into each methodology, as understood by scholars in the respective field. The disconnects and gaps uncovered from this analysis suggest opportunities for basic researchers to pursue lines of inquiry that will make a difference for applied decision research.

Table 8.1 summarizes central themes for ready comparison. With respect to the first, note that the topic and product of these two kinds of applied research are distinct. So, at the methodological level, they really are non-competing. Furthermore, CTA comprises a front-end analysis that can provide useful supporting information pertaining to the suitability of DA to the problem at hand, as well as the problem-structuring step of a DA. Combining CTA and DA in this way may seem odd to those who think of CTA as sanctifying intuition, but it is important to keep in mind that some form of analysis is often found among the expert decision-makers' strategies.

A related point of confusion is the idea that CTA is best used for short-term tactical types of problems, whereas DA is best suited for long-range strategic issues. The sponsors for CTA and DA research may cluster toward the tactical or the strategic, creating this impression, but the methods themselves are suited to both tactical and strategic problems. The one possible exception is the use of extensive DA in the midst of tactical decision-making. The time constraints do render application of DA unsuitable for the immediate situation in such cases. However, it is possible to use DA to evaluate policies in research studies outside of those situations, and then employ the results (e.g., via training) in the situations themselves.

Table 8.1 Decision analysis and cognitive task analysis compared

Theme	DA	CTA
Analysis		
Topic	Specific decision	Decision-maker cognition
Product	Recommended course of action	Decision requirements, recommended design
Quality and achievement	Consistency with axioms, extensive analysis	Bootstrapping expertise
Role of the decision-maker	Render probabilities and utilities, judge usability and acceptability of solution	Research participant, collaborator in design, judge usability and acceptability of solution
Role of the psychologist	Elicit probabilities and utilities	Lead research project

E1

The conceptions of decision quality of the two approaches do appear to conflict rather directly. Furthermore, neither conception seems entirely satisfactory. The classic arguments used to support the need for consistency with the axioms, such as avoidance of becoming a money pump, are fairly weak and ad hoc from an applied standpoint. That is, it is far from clear that becoming a money pump is a genuine risk that plagues decision-makers in the real world. As important as the axioms themselves are for the theoretical development of analytical methods (e.g., statistical methods), consistency with them is an insufficient standard of decision performance (cf. Yates 1982). In addition, use of extensive quantitative analysis is a process, not a performance metric. And the link to performance is far from clear (Klein *et al*. 1995; Gigerenzer *et al*. 1999; Gigerenzer 2005).

From the CTA standpoint, significant research and thought are required to determine how the concept of expert performance can be widely appreciated as a gold standard, given the process–performance paradox. The preliminary clarifications made here regarding the concept of "bootstrapping expertise" and examination of the degree to which solutions meet the specific needs of decision-makers allow for some potential points of departure of such research. Also, at this point, the expert model is better thought of as providing a baseline reference to be accounted for in improvement efforts, rather than an ideal standard. NDM does not claim that the expert has reached the pinnacle of good decision-making. This, of course, implies that both approaches are lacking in generally acceptable ideal standards.

A related question is whether experts are, in fact, any good at what they do (Smith *et al*. 2004). A number of studies have compared experts with novices and decision aids on experimentally well-defined tasks that permit convenient quantitative comparisons, finding the experts to be wanting. From our standpoint, "what they do" is a critical issue. This issue has been discussed extensively (Rohrbaugh & Shanteau 1999; Phillips *et al*. 2004). For example, the value is unclear of comparisons between individual expert performance and novices (or decision aids) on experimental tasks generated by relatively uninformed experimenters. That is, the task constraints often appear such that the experiments are not capturing the judgments and processes that make up expertise. For example, Westen and Weinberger (2005) argued that the research showing the limitations of experts does not provide a reasonable assessment. If clinicians are asked a question such as whether a specific client will commit a violent act, the clinician is not likely to give an accurate appraisal simply because clinicians have very few patients who engage in violence. Thus, the clinicians have little expertise at making this type of judgment. If skilled clinicians are asked about client features that they frequently encounter, their accuracy rates are much better. Sieck (2003) made a similar observation regarding differences in calibration between experts who render probability judgments as part of their jobs and those who do not. It is worth pointing out that the mismatch between actual expertise and elicited judgments is of particular concern for the application of DA, where quantity judgments are required in a form that is often incompatible with the decision-maker's expertise.

In sum, a broader perspective is required to get at the nature of decision quality. The discussions provided here, as well as recent research by Yates and his colleagues, may provide serious beginnings for addressing these issues (Yates 2003; Yates *et al*. 2003; Hoffman & Yates 2005).

Our third theme, concerning the role of the expert decision-maker, has non-trivial implications for the acceptability and implementation of the final recommendations. From

E1

the NDM perspective, it is essential to include the decision-makers at the front end, and understand the issues from their perspective. Even if some of the extreme critics of expertise are correct (which we seriously doubt), the results of getting decision support tools designed according to DA principles successfully integrated and implemented into practice have been less than satisfactory (Corey & Merenstein 1987; Ashton 1991). Although it is common to berate the decision-makers and trivialize their experience in such cases, such a position does not serve the common goal of improving decision-making. Decision support solutions based on designs that take the challenges of the decision-makers into account are likely to be much more useable and acceptable to those experts. And applied research has to take usability and acceptability of the recommendations into account, in addition to scientific validity of the process.

Our intent here is to get beyond confrontations between DA and CTA for the two methodologies are not necessarily opposed. They are each attempting to meet different objectives. In that frame, we believe that psychologists who can appreciate aspects of both approaches can work on ways to use CTA field research methods to improve DA. Likewise, DA can be used to bolster CTA.

Thus far, decision analysts have concentrated on finding stronger techniques and better ways to do the math. However, the big issues with DA have to do with the people, not the math. The practice of DA can be improved by understanding the broader challenges of knowledge elicitation and problem structuring. DA will improve by devising better methods for gaining acceptance within organizations, and by becoming more sensitive to the work context and organizational issues that have been barriers to adoption. Methods of CTA may be helpful in taking these steps. On the other hand, CTA researchers have focused their efforts on developing methods to investigate expert decision-making in natural settings, with the aim of uncovering a wide range of solutions that meet the specific needs of decision-makers. Little work has been done, however, to generate more external evaluations of recommendations. To achieve broader acceptance of the approach, DA principles could be used to validate and quantify the benefits of CTA. Furthermore, the graphical and formal representations used in DA may provide useful constructs for representing expert decision processes in the data analysis phase of CTA. At this juncture, the investigation of combined CTA–DA approaches for improving decision-making is surely warranted.

AUTHOR NOTE

The authors thank Frank Durso and three reviewers for their patience and comments on an earlier version of this chapter.

REFERENCES

Ashton, R. (1991). Pressure and performance in accounting decision settings: paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research*, *28*, 148–186.

Beccue, P. (2001). Choosing a development strategy for a new product at Amgen. *Interfaces*, *31*(5), 62–64.

E1

Benson, P. G., Curley, S. P. & Smith, G. F. (1995). Belief assessment: an underdeveloped phase of probability elicitation. *Management Science*, *41*(10), 1639–1653.

Berntsen, D. & Thomsen, D. K. (2005). Personal memories for remote historical events: accuracy and clarity of flashbulb memories related to World War II. *Journal of Experimental Psychology: General*, *134*(2), 242–257.

Birnbaum, M. H. (2004). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes*, *95*, 40–65.

7 Birnbaum, M. H. (in press). Three new tests of independence that differentiate models of risky decision making. *Management Science*, ••, ••–••.

Bodily, S. E. & Allen, M. S. (1999). A dialogue process for choosing value-creating strategies. *Interfaces*, *29*(6), 16–28.

Browne, G. J., Curley, S. P. & Benson, P. G. (1997). Evoking information in probability assessment: knowledge maps and reasoning-based directed questions. *Management Science*, *43*(1), 1–14.

Cipollini, K .A., Maruyama, A. L. & Zimmerman, C. L. (2005). Planning for restoration: a decision analysis approach to prioritization. *Restoration Ecology*, *13*(3), 460–470.

8 Clemen, R. T. & Kwit, R. C. (2001). The value of decision analysis at Eastman Kodak Company, 1990–1999. *Interfaces*, *31*(5), 74–92.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, *41*, 801–849.

Corey, G. A. & Merenstein, J. H. (1987). Applying the acute ischemic heart disease predictive instrument. *The Journal of Family Practice*, *25*, 127–133.

Crandall, B., Klein, G. & Hoffman, R. R. (2006). *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. Boston, MA: MIT Press.

Dawes, R. M. & Corigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*, 95–106.

Dawson, N. V., Connors, A. F. Jr., Speroff, T. *et al*. (1993). Hemodynamic assessment in managing the critically ill: is physician confidence warranted? *Medical Decision Making*, *13*, 258–266.

Edwards, W. (1998). Hailfinder: tools for and experiences with Bayesian normative modeling. *American Psychologist*, *53*(4), 416–428.

Edwards, W. & Fasolo, B. (2001). Decision technology. *Annual Review of Psychology*, *52*, 581–606.

Edwards, W., Kiss, I., Majone, G. & Toda, M. (1984). What constitutes "a good decision?" *Acta Psychologica*, *56*, 5–27.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*, 327–358.

Geiselman, R. E. & Fisher, R. P. (1997). Ten years of cognitive interviewing. In D. G. P. F. G. Conrad (ed.), *Intersections in Basic and Applied Memory Research*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gigerenzer, G. (2005). Is the mind irrational or ecologically rational? In F. Parisi & V. L. Smith (eds.), *The Law and Economics of Irrational Behavior* (pp. 37–67). Stanford, CA: Stanford University Press.

Gigerenzer, G., Todd, P. M. & Group, T. A. R. (1999). *Simple Heuristics that Make Us Smart*. New York: Oxford University Press.

Hammond, J. S., Keeney, R. L. & Raiffa, H. (1999). *Smart Choices: A Practical Guide to Making Better Decisions*. Boston, MA: Harvard Business School Press.

Hammond, K. R. & Adelman, L. (1976). Science, values, and human judgment. *Science*, *194*, 389–396.

Hoffman, R. R., Crandall, B. W. & Shadbolt, N. R. (1998). Use of the critical decision method to elicit expert knowledge: a case study in cognitive task analysis methodology. *Human Factors*, *40*(2), 254–276.

Hoffman, R. R. & Yates, J. F. (2005). Decision (?) Making (?). *IEEE Intelligent Systems*, *20*(4), 76–83.

Hollnagel, E. (ed.) (2002). *Handbook of Cognitive Task Design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Howard, R. A. (1989). Knowledge maps. *Management Science*, *35*, 903–922.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press. E1

Hutton, R. J. B. & Klein, G. (1999). Expert decision making. *Systems Engineering*, *2*(1), 32–45.

Hutton, R. J. B. & Miller, T. E. (2003). Decision-Centered Design: leveraging cognitive task analysis in design. In E. Hollnagel (ed.), *Handbook of Cognitive Task Design*. Mahwah, NJ: Lawrence
9 Erlbaum Associates.

Johnson, E. (2003). Analyzing the development strategy for apimoxin. *Interfaces*, *33*(3), 57–59.

Klein, G. (1998). *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press.

Klein, G., McCloskey, M. J., Pliske, R. M. & Schmitt, J. (1997). Decision skills training. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 182–185).
10

Klein, G., Phillips, J. K., Rall, E. L. & Peluso, D. A. (in press). A data/frame theory of sensemaking. In R. R. Hoffman (ed.), *Expertise Out of Context: Proceedings of the 6th International Conference on Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, G., Wolf, S., Militello, L. G. & Zsambok, C. E. (1995). Characteristics of skilled option generation in chess. *Organizational Behavior and Human Decision Processes*, *62*(1), 63–69.

Klein, G. A. (1989). Recognition-primed decisions. In W. B. Rouse (ed.), *Advances in Man-Machine Systems Research*. Greenwich, CT: JAI Press.

Klein, G. A., Calderwood, R. & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(3), 462–472.

Kusnic, M. W. & Owen, D. (1992). The unifying vision process: value beyond traditional decision analysis in multiple-decision-maker-environments. *Interfaces*, *22*(6), 150–166.

Lee, P. M. (2004). *Bayesian Statistics: An Introduction* (3rd edn). London: Oxford University Press.

Lichtenstein, S. & Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance*, *26*, 149–171.

Mellers, B. A. (2000). Choice and the relative pleasure of consequences. *Psychological Bulletin*, *126*, 910–924.

Mellers, B. A., Schwartz, A. & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General*, *128*, 1–14.

Militello, L. G. & Hutton, R. J. B. (1998). Applied Cognitive Task Analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, *41*(11), 1618–1641.

Nisbett, R. E. & Wilson, T. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Peterson, M. F., Miranda, S. M., Smith, P. B. & Haskell, V. M. (2003). The sociocultural contexts of decision making in organizations. In S. L. Schneider & J. Shanteau (eds.), *Emerging Perspectives on Judgment and Decision Research* (pp. 512–555). Cambridge: Cambridge University Press.

Phillips, J. K., Klein, G. & Sieck, W. R. (2004). Expertise in judgment and decision making: a case for training intuitive decision skills. In D. J. Koehler & N. Harvey (eds.), *Blackwell Handbook of Judgment & Decision Making* (pp. 297–315). Malden, MA: Blackwell.

Pliske, R. M., Crandall, B. & Klein, G. (2004). Competence in weather forecasting. In K. Smith, J. Shanteau & P. Johnson (eds.), *Psychological Investigations of Competence in Decision Making* (pp. 40–70). Cambridge: Cambridge University Press.

Pollock, S. M. & Chen, K. (1986). Strive to conquer the black stink: decision analysis in the People's Republic of China. *Interfaces*, *16*(2), 31–37.

Rohrbaugh, C. C. & Shanteau, J. (1999). Context, process, and experience: research on applied judgment and decision making. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay & M. T. H. Chi (eds.), *Handbook of Applied Cognition* (pp. 115–139). Chichester: John Wiley & Sons.

Schraagen, J. M., Chipman, S. F. & Shalin, V. L. (eds.) (2000). *Cognitive Task Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schraagen, J. M., Chipman, S. F. & Shute, V. J. (2000). State-of-the-art review of cognitive task analysis techniques. In J. M. Schraagen, S. F. Chipman & V. L. Shalin (eds.), *Cognitive Task Analysis* (pp. 467–487). Mahwah, NJ: Lawrence Erlbaum Associates.

Shanteau, J., Weiss, D. J., Thomas, R. P. & Pounds, J. (2003). How can you tell if someone is an expert? Performance-based assessment of expertise. In S. L. Schneider & J. Shanteau (eds.),
E1 *Emerging Perspectives on Judgment and Decision Research* (pp. 620–639). Cambridge: Cambridge University Press.

Sieck, W. R. (2003). Effects of choice and relative frequency elicitation on overconfidence: further tests of an exemplar-retrieval model. *Journal of Behavioral Decision Making*, *16*, 127–145.
Sieck, W. R. & Arkes, H. R. (2005). The recalcitrance of overconfidence and its contribution to decision aid neglect. *Journal of Behavioral Decision Making*, *18*, 29–53.
Smith, K., Shanteau, J. & Johnson, P. (2004). *Psychological Investigations of Competence in Decision-making*. Cambridge: Cambridge University Press.
Staszewski, J. (2004). Models of human expertise as blueprints for cognitive engineering: applications to landmine detection. Paper presented at the Human Factors and Ergonomics Society.
Weick, K. E. (1995). *Sensemaking in Organizations*. Newbury Park, CA: Sage.
Westen, D. & Weinberger, J. (2005). In praise of clinical judgment: Meehl's forgotten legacy. *Journal of Clinical Psychology*, *61*(10), 1257–1276.
von Winterfeldt, D. & Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge: Cambridge University Press.
Wright, G. & Wisudha, A. (1982). Distribution of probability assessments for alamanc and future event questions. *Scandinavian Journal of Psychology*, *23*(3), 219–224.
Yates, J. F. (1982). External correspondence: decompositions of the mean probability score. *Organizational Behavior and Human Decision Processes*, *30*, 132–156.
Yates, J. F. (2003). *Decision Management: How to Assure Better Decisions in Your Company*. San Francisco, CA: Jossey-Bass.
Yates, J. F., Veinott, E. S. & Patalano, A. L. (2003). Hard decisions, bad decisions: on decision quality and decision aiding. In S. L. Schneider & J. Shanteau (eds.), *Emerging Perspectives on Judgment and Decision Research* (pp. 13–63). New York: Cambridge University Press.
Zsambok, C. E. & Klein, G. (eds.) (1997). *Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates.

E1

AUTHOR QUERY-FORM

Dear Author,

During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof.

Many thanks for your assistance.

Query References	Query	Remarks
1.	Au: **Clemen 1996** has not been found in the reference list.	
2.	Au: **Clemen 1996** has not been found in the reference list.	
3.	Au: **Goodwin and Wright 1998** has not been found in the reference list.	
4.	Au: **Edwards and Barron 1994** has not been found in the reference list.	
5.	Au: **Klein *et al.* 1985** has not been found in the reference list.	
6.	Au: **Hutton *et al.* (2003)** has not been found in the reference list.	
7.	Au: Please supply the volume number and the page range if available.	
8.	Au: **Clemen and Kwit 2001** has not been found in the text. Please indicate where it should be cited or delete this reference.	
9.	Au: **Hutton and Miller 2003** has not been found in the text. Please indicate where it should be cited or delete this reference.	
10.	Au: Please supply the publisher.	