

Focus Article: Taking Stock of Naturalistic Decision Making

RAANAN LIPSHITZ^{1*}, GARY KLEIN², JUDITH ORASANU³
and EDUARDO SALAS⁴

¹*University of Haifa, Israel*

²*Klein Associates, USA*

³*NASA Ames, USA*

⁴*University of Central Florida, USA*

ABSTRACT

We review the progress of naturalistic decision making (NDM) in the decade since the first conference on the subject in 1989. After setting out a brief history of NDM we identify its essential characteristics and consider five of its main contributions: recognition-primed decisions, coping with uncertainty, team decision making, decision errors, and methodology. NDM helped identify important areas of inquiry previously neglected (e.g. the use of expertise in sizing up situations and generating options), it introduced new models, conceptualizations, and methods, and recruited applied investigators into the field. Above all, NDM contributed a new perspective on how decisions (broadly defined as committing oneself to a certain course of action) are made. NDM still faces significant challenges, including improvement of the quantity and rigor of its empirical research, and confirming the validity of its prescriptive models. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS naturalistic decision making; recognition-primed decisions; coping with uncertainty; team decision making; decision errors; decision training; research methodology

The study of decision making is studded by three-letter acronyms designating sub-disciplines which evolved partly as extensions of preceding sub-disciplines, and partly as a reaction to them: the once-popular CDM (Classical Decision Making), BDT (Behavioral Decision Theory), JDM (Judgment and Decision Making), ODM (Organizational Decision Making), and, most recently, NDM (Naturalistic Decision Making). The emergence of each sub-discipline can be conveniently traced to the publication of books or papers signifying the time at which theory and research pursued more or less in isolation gathered sufficient mass and coherence to attract wider attention. CDM can be traced to Bernoulli (1738) and, more recently, to Savage (1954) and von Neumann and Morgenstern (1944). BDT and JDM have their origins in Edwards (1954) and Meehl (1954).

*Correspondence to: Professor Raanan Lipshitz, Department of Psychology, University of Haifa, Haifa, Israel, 31905.
E-mail: raanan@psy.haifa.ac.il

ODM can be traced to Simon (1957), March and Simon (1958), and Cyert and March (1963). Finally, NDM goes back to Klein *et al.* (1993). A decade has now passed since the conference that produced the last-named volume, a sufficiently long period to take stock of NDM: its essential characteristics, strengths, weaknesses, and future prospects. After drawing a historical sketch of NDM, we present its essential characteristics and examine critiques of its theoretical bases, methodology, and contributions, focusing on five areas: recognition-primed decisions, coping with uncertainty, decision errors, team decision making, and decision aiding and training. We close the paper by drawing some conjectures regarding the future directions of NDM.

A BRIEF HISTORY OF NDM

The NDM framework was initiated in 1989 in a conference in Dayton, Ohio, sponsored by the Army Research Institute. The conference enabled some 30 behavioral scientists working in academic and non-academic institutions to discover that they shared many common themes, regardless of domain. One theme was the importance of time pressure, uncertainty, ill-defined goals, high personal stakes, and other complexities that characterize decision making in real-world settings. Although these factors were difficult to replicate in the laboratory, they needed to be understood (Orasanu and Connolly, 1993). A second theme was the importance of studying people who had some degree of expertise; novices were never used in the study of the type of high-stakes tasks that were of interest (Pruitt *et al.*, 1997). A third theme was that the way people sized up situations seemed more critical than the way they selected between courses of action (Klein, 1993).

In the past ten years there has been an increasing amount of interest in NDM. The 1989 conference (Klein *et al.*, 1993) was followed by a second conference (Zsambok and Klein, 1997) held in 1994 and attended by approximately 100 researchers. A third NDM conference was held in Aberdeen, Scotland, in 1996 (Flin *et al.*, 1997), and a fourth conference was held in Warrenton, Virginia, in 1998 (Salas and Klein, in press). In addition to the edited volumes emerging from each conference, Flin (1996) has written about the issues facing critical incident managers, Klein (1998) has described the work of his research group, and Cannon-Bowers and Salas (1998) edited a book describing the research program sponsored by the US Navy in the aftermath of the *Vincennes* incident. Finally, Beach (1997) surveyed NDM from the vantage point of his own work on Image Theory, a model that is aligned with the NDM framework. In addition to these publications, the Human Factors and Ergonomics Society established a technical group in 1995, called 'cognitive engineering and decision making', partly as an outlet for research and development along the lines of NDM. As of 1998 there were more than 500 members, making it one of the largest technical groups in the Society.

ESSENTIALS OF NATURALISTIC DECISION MAKING

NDM is an attempt to understand how people make decisions in real-world contexts that are meaningful and familiar to them. Fulfilling this 'mission' produced research marked by five essential characteristics: *proficient decision makers, situation-action matching decision rules, context-bound informal modeling, process orientation, and empirical-based prescription*. These particular characteristics were derived by locating the place of NDM in the study of decision making based on Rasmussen's (1997) observation that

In several human sciences [including decision research], a trend is found in modeling behavior: Efforts are moving from normative models of rational behavior [e.g., CDM], through efforts to model the observed rational behavior by means of models of the deviation from rational [e.g., JDM], toward focus on representing directly the actually observed behavior [e.g., NDM], and ultimately to efforts to model behavior generating mechanisms [i.e., models of system constraints, opportunities and criteria, e.g., ODM] (p. 75, material in brackets added by us).

Granted that reconstruction of history inevitably fineses subtle twists and turns in the actual turn of events, Rasmussen's sequence does fairly well (ODM in fact preceded NDM and fits into Rasmussen's sequence only in terms of the move from individual to system-wide models). Our historical perspective suggests that one way of deriving the essential characteristics of NDM is to examine its differences from CDM, the preceding phase in Rasmussen's sequence.

The essential characteristics of CDM were (1) *choice* (conceptualizing decision making as choosing among concurrently available alternatives, e.g. Dawes, 1988; Hogarth, 1987), (2) *input–output orientation* (focusing on predicting which alternative will, or should be, chosen given a decision maker's preferences: Funder, 1987), (3) *comprehensiveness* (conceptualizing decision making as a deliberate and analytic process that requires a relatively thorough information search (Beach and Mitchell, 1978, Payne *et al.*, 1990), particularly for optimal performance (Gigerenzer and Todd, 1999; Grandori, 1984), and (4) *formalism* (the development of abstract, context-free models amenable to quantitative testing, e.g. Coombs *et al.*, 1971). The history of decision research consists of the gradual replacement of these characteristics, beginning with doubts regarding their effects on the descriptive validity of CDM and culminating in the replacement of all four by other characteristics for descriptive as well as prescriptive purposes in NDM.

Doubts regarding the validity of the rational choice model as a valid description of human decision making probably preceded the work of Simon and his associates at Carnegie-Mellon University. However, their contribution was seminal because it went beyond just pointing out that the informational requirements (i.e. comprehensiveness) entailed in the model exceed limited human cognitive capacities. Through the concept of bounded rationality, which points to attention as the scarce resource in human decision making, Simon *et al.* showed that people's systematic deviations from the rational choice model make sense from an adaptive perspective: under bounded rationality thoroughgoing information processing is exhausting, and potentially futile. A second, and just as important though less publicized, proposition of the Carnegie School was an attack on the prescriptive validity of the Rational Choice model. As Simon (1978) suggested, real-world problems are typically loosely coupled, allowing decision makers with bounded rationality to attend to them effectively in a sequential fashion. Thus, effective adaptation does not require comprehensive analysis. Instead, all that are required are a modest intellectual capacity, an ability to detect and prioritize problems, and the ability to learn from experience.

JDM/BDT further undermined the descriptive validity of CDM, showing that people tend to deviate systematically from the rational choice model even when presented with relatively simple tasks which do not severely tax bounded rationality (Kahneman *et al.*, 1982). However, JDM/BDT retained the essential characteristics of CDM and adhered to its normative models as standards for evaluating decision quality. Thus, Elimination by Aspects (Tversky, 1972), Prospect Theory (Kahneman and Tversky, 1979), and Einhorn and Hogarth's (1986) Ambiguity Model, as three representative examples, are all formal choice models that describe which alternative is chosen from an available set of alternatives based on different comparison schemes. In addition, JDM/BDT texts prescribe Multi-Attribute Utility (MAU)-like and Subjective Expected Utility (SEU)-like procedures (Russo and Schoemaker, 1987) and 'de-biasing' procedures for correcting deviations from these models (Fischhoff, 1982).

Going beyond JDM/BDT's criticism of CDM, NDM replaced all the four essential characteristics of the latter identified above. Comprehensive choice was replaced by matching, input–output orientation was replaced by process orientation, and context-free formal modeling was replaced by context-bound informal modeling. It is fair to say that these characteristics followed once researchers within the NDM framework embarked on the construction of descriptive models of proficient decision makers in natural contexts without relying on normative choice models as starting points. Following the emphasis on bounded rationality of the Carnegie School, NDM places the human (and hence boundedly rational) proficient decision maker at its center of interest and as its basis for prescription.

Proficient decision makers

In the decade since the first NDM conference in 1989, the definition of NDM has changed. It is marked by a shift in the relative emphasis placed on expertise and features of field settings in which decisions are made. The original definition proposed by Orasanu and Connolly (1993) emphasized the shaping features of the *contexts* in which many decisions of interest were made: ill-structured problems, uncertain, dynamic environments, shifting, ill-defined, or competing goals, multiple event-feedback loops, time constraints, high stakes, multiple players, and organizational settings. Expertise was included as a secondary factor.

By the time of the second NDM conference, an alternative definition had emerged. Zsambok (1997, p. 4) distinguishes NDM in terms of the decision maker, positing that 'NDM is the way people use their experience to make decisions in field settings'.

Pruitt *et al.* (1997) went one step further, and concluded that the primary factor defining NDM studies is expertise:

[I]t is possible to answer the question of knowing an NDM study . . . by looking at how the study handles the subject's prior experience . . . Does the study treat prior experience as a nuisance variable (one to be controlled, counterbalanced, or otherwise ignored) or does it view this variable as the focus of inquiry? We would argue that CDM [BDT, and JDM] do the former and NDM does the latter . . . [T]he strength of NDM is its emphasis on experience and knowledge which already is present in the subject. Looking back at the short definition of Zsambok [above]. . . we believe that the inclusion of 'in field settings' is only secondary (pp. 37–38).

Still, we cannot ignore the influence of field settings because they establish the eliciting conditions for making decisions and shape decisions through their constraints and affordances. 'Expertise' is about these field settings.

Granted that NDM is concerned with proficient decision makers, namely people with relevant experience or knowledge in the decision-making domain who rely on their experience directly, the remaining four essential characteristics of NDM follow:

- *Process orientation:* In contrast to input–output orientation, NDM models do not attempt to predict which option will be implemented, but describe the cognitive processes of proficient decision makers. This difference in orientation has important implications for validation (Funder, 1987). To be valid, NDM models have to describe what information decision makers actually seek, how they interpret it, and which decision rules they actually use. This is another reason why NDM models tend not to be formal, and especially not abstract. Initial studies of the process by which experts make decisions have yielded the next distinguishing feature.
- *Situation–action matching decision rules:* Matching is a generic label for decisions with the basic structure of 'Do A because it is appropriate for situation S' (Lipshitz, 1994). The study of proficient decision makers leads to modeling decision making as matching rather than choice. Numerous studies have consistently shown that proficient decision makers typically make decisions by various forms of matching and not by concurrent choice (i.e. 'Do A because it has superior outcomes to its alternatives'). For example, Newell and Simon (1972) modeled the decision making of expert chess players as a system of nested matching rules, March (1982) suggested that decisions in organizational contexts follow the logic of obligation (which dictates what is appropriate for persons in specific roles to do in specific situations), and Carroll and Payne found that parole officers make decisions by matching candidate features to different prototypes of offenders (Carroll, 1980). Matching differs from concurrent choice in three respects. (1) Options are evaluated sequentially one at a time. Evidence exists that even when presented with several options, decision makers quickly screen most of them by comparing them against a standard, rather than with one

another, and then focus on one, or at most two, options, which are compared (Beach, 1993; Montgomery, 1988). (2) Options are selected or rejected based on their compatibility with the situation (Endsley, 1997; Klein, 1998; Pennington and Hastie, 1993), or the decision maker's values (Beach, 1990) rather than on their relative merits. (3) The process of matching may be analytic but more often it relies on pattern matching and informal reasoning (Cohen *et al.*, 1996; Klein, 1998; Lipshitz, 1993; Pennington and Hastie, 1986). Some of these variations are discussed in more detail in the section on recognition-primed decisions below.

- *Context-bound informal modeling:* As noted above, proficient decision making is driven by experience-tied knowledge. This puts a limit on the utility of abstract formal models for two reasons: (1) expert knowledge is domain- and context-specific (Ericsson and Lehman, 1996; Smith, 1997); (2) decision makers are sensitive to semantic as well as syntactic content (Wagenaar *et al.*, 1988; Searle, 1995). For this reason NDM models depict what information decision makers actually attend to and which arguments they actually use, particularly if they are designed for applied purposes (e.g. Cohen and Freeman, 1997; Crandall and Getchell-Reiter, 1993).
- *Empirical-based prescription:* JDM and BDT derive prescriptive models from normative models which stand upon explicit formal proofs of optimization believed to be independent of the descriptive validity of these models. This means that 'ought' can be divorced from 'is', namely that solutions can be prescribed irrespective of the intended recipient's ability to perform them. NDM researchers believe that 'ought' cannot be divorced from 'is': prescriptions which are optimal in some formal sense but which cannot be implemented are worthless. This leads to empirical-based prescription, namely deriving prescriptions from descriptive models of expert performance. The goal of empirical-based prescription, then, is to improve feasible decision makers' characteristic modes of making decisions (e.g. sequential single-option evaluation), rather than replacing them altogether, by basing prescription on demonstrations of feasible expert performance.

Empirical-based prescription is consistent with the observation, noted in the section on context-bound modeling, that decision makers in natural settings use situated content-driven cognitive processes to solve domain-specific problems by taking concrete actions (Klein *et al.*, 1993). This implies that empirical-based prescription is valid only under conditions that permit the development of true expertise (e.g. the availability of repetitive tasks and valid feedback; Shanteau, 1992). In addition, it implies three tradeoffs with clear methodological implications. First, there is a tradeoff between the generality of prescriptive models and their applicability. Since general models are by definition non-specific, they are likely to be misinterpreted (Reason, 1990) or fail to match critical requirements peculiar to the problem at hand (Smith, 1997). Second, structural models that specify the general functional relationships among variables, and which are tested, however validly, in laboratory studies, do not provide information on how to change *X* in order to achieve change in *Y*. For example, a model which specifies that decision effectiveness is a function of the optimality of information search is not informative as to how information search can be optimized in a particular task situation. Thus, there is a tradeoff between the theoretical value of models and research methods and the 'actionability' of the knowledge that they provide (i.e. its usefulness as guide for action; Argyris, 1993). Finally, while formal analytic models can yield optimal solutions with great precision and rigor, they can also be inefficient owing to the cognitive effort which they require (Beach and Mitchell, 1978), their poor compatibility with decision makers' problems (Humphreys and Berkeley, 1985; Smith, 1997) and the non-analytic cognitive processes which decision makers typically use (Hammond, 1993).

Although a number of models fall within the NDM framework (Lipshitz, 1993), it is fair to say that the RPD model (Klein, 1993, 1998) can serve as the prototypical NDM model. The next section goes into some detail on the RPD model to illustrate the essential characteristics of this approach and how a naturalistic account is used in research.

RECOGNITION-PRIMED DECISION MAKING

The RPD model was developed on the basis of cognitive task analyses of firefighters (Klein *et al.*, 1989). The initial research was designed to better understand how experienced commanders could handle time pressure and uncertainty. The purpose of this research was not to challenge traditional decision making but to conduct a descriptive inquiry. The investigators hypothesized that under time pressure, commanders would not be able to generate a large set of response options, but would be likely to fall back on a simple comparison between a favored option and a comparison option. Probe question-based interviews were conducted with more than 30 firefighters with an average of 23 years of experience, to obtain retrospective data about 156 highly challenging incidents. The data suggested that in most cases the commanders were not comparing any options. They were typically carrying out the first course of action they identified. This raised two questions: how could the commanders rely on the first option they considered, and how could they evaluate a single option, without the decision maker comparing it to any others?

The model was formulated by synthesizing the descriptions provided by the commanders themselves. In its current form, the RPD model has three variations. In the simplest variation of the model, a decision maker sizes up a situation and responds with the initial option identified. The hypothesis is that skilled decision makers can usually generate a feasible course of action as the first one they consider, which answers the first question above, about how commanders could rely on the first option they considered. In this variation, experience provides prototypes or functional categories. This is different from retrieving analogues, although some analogical reasoning may be involved. Skilled decision makers perceive situations as typical cases where certain types of actions are typically appropriate, and are usually successful.

The second variation (which emerged from a similar type of study with commanders of AEGIS cruisers, conducted by Kaempf *et al.*, 1996), describes what happens if the situation is not clear. Here, the skilled decision maker will often rely on a story-building strategy to mentally simulate the events leading up to the observed features of the situation. This type of strategy has been described by Pennington and Hastie (1993) and by Klein and Crandall (1995).

The third variation describes how decision makers can evaluate a course of action without comparing it to others, which is the second question raised above. The evaluation is conducted by mentally simulating the course of action, to see if it will work, and to look for unintended consequences that might be unacceptable. De Groot (1965) referred to this strategy as progressive deepening.

These three variations depend heavily on expertise. In the first variation, expertise provides a sense of typicality that allows decision makers to quickly categorize situations and to recognize how to react as an aspect of the categorization. In the second variation, expertise is needed to construct the mental models needed to find one explanation more plausible than another. In the third variation, expertise is defined as an ability to mentally simulate a course of action in a situation, and anticipate how it will play out.

The three variations explain how decision makers can handle the constraints and stressors often found in field settings. Under extreme time pressure, the first variation will result in reasonable reactions without the need to perform any deliberations or analyses. Under uncertainty, the second variation describes how the plausibility of alternative stories can help a decision maker choose an interpretation, and categorize a situation. Under shifting conditions, the decision maker is prepared to react quickly, without having to re-do analyses. When faced with ill-defined goals, the decision maker is not stymied because the RPD model is aimed at working forwards, from existing conditions, rather than backwards, from goal states. Patel and Groen (1986) and Larkin *et al.* (1980) have shown that people with greater expertise are more likely to use forward-chained reasoning, whereas novices and intermediate subjects usually rely on backward-chained reasoning.

The initial findings of the research with firefighters have been replicated several times, by different research teams (see Klein, 1998, for a review). These studies have been conducted with naval surface ship

commanders, tank platoon leaders, wildfire as well as urban fire commanders, design engineers, offshore oil installation managers, infantry officers, and commercial aviation pilots. The data have been coded for different types of decision strategies, and the RPD strategy has usually been shown to be the most common, representing 80–95% of the cases. Only with very inexperienced decision makers does the proportion fall below 50%.

Klein (1998) has described some of the boundary conditions for the RPD model. It appears to hold when there is reasonable experience to draw on, when the decision maker is under time pressure, and when there is uncertainty and/or ill-defined goals. The RPD strategies are less likely to be used with highly combinatorial problems, in situations where justifications are required, and in cases where the views of different stakeholders have to be taken into account.

The RPD model has been used to generate testable hypotheses. One confirmed prediction was that extreme time pressure would have a minimal effect on chess masters, as compared with mediocre players. Calderwood *et al.* (1988) showed that the proportion of poor moves was basically the same for chess masters playing actual games, regardless of whether the games were played using regulation time (40 moves in 90 minutes) or blitz conditions (5 minutes total for the game). The mediocre players showed a sharp increase in poor moves under time pressure. A second prediction is that skilled chess players could generate a reasonable move as the very first one they considered. Klein *et al.* (1995) obtained think-aloud protocols from both mediocre and skilled chess players working on a series of difficult chess problems. Grandmaster ratings of these positions showed that only 1/6 of the legal moves were considered adequate. The finding from the think-aloud protocols was that 4/6 of the actual first moves considered were adequate, according to the grandmaster criteria. Clearly, the subjects were not generating courses of action by randomly selecting from the pool of legal options. They were using their expertise to generate a good move as the first one they considered. We are not aware of any decision theories that predict the opposite, that people randomly generate options, so this is not a critical experiment. Nevertheless, the findings do contribute to our understanding of how expertise can influence decision-making strategies. The result has implications for prescriptions such as multi-attribute utility analysis. If a moderately experienced person can generate a workable option as the first one considered, there may be reduced incentives and benefits from generating and evaluating additional courses of action.

COPING WITH UNCERTAINTY

The attributes that Orasanu and Connolly (1993: see above) identified as characteristics of natural decision making can be clearly linked to the uncertainty and stress that accompany the making of consequential decisions in naturalistic settings (the exception being ‘organizational settings’). The RPD model accounts for the fact that proficient decision makers perform reasonably (and at times exceptionally) well under these conditions by their effective use of pattern matching, forward-directed reasoning, and storytelling. Two NDM models which focus on how decision makers cope with uncertainty, the RAWFS heuristic (Lipshitz, 1997a; Lipshitz and Strauss, 1997) and the Recognition/Meta-cognition (R/M) model (Cohen *et al.*, 1998), elaborate these and suggest additional strategies.

The RAWFS heuristic addresses three questions: (1) How do decision makers conceptualize uncertainty? (2) How do they cope with uncertainty? (3) Are there systematic relationships between different conceptualizations of uncertainty and methods of coping? Lipshitz and Strauss began by defining uncertainty in the context of action as ‘a sense of doubt that blocks or delays action’, an inclusive definition which is consistent with Dewey (1933), and accommodates the numerous definitions of uncertainty in the JDM/BDT as well as ODM literatures. The definition is also supported by findings that people evaluate ‘decisions’ as ‘certain’, ‘active’, ‘quick’, and ‘strong’, and uncertainty as ‘passive’, ‘slow’, and ‘weak’, on a set of semantic scales (Teigen, 1996). Using this definition, Lipshitz and Strauss identified three principal forms of uncertainty in

retrospective reports of decision making under uncertainty: inadequate understanding (a sense of having an insufficiently coherent situation awareness), lack of information (a sense of having incomplete, ambiguous, or unreliable information), and conflicted alternatives (a sense that available alternatives are insufficiently differentiated). (Orasanu and Fischer, 1997, proposed a similar conceptualization based on observations of commercial airplane crews.)

In addition, Lipshitz and Strauss found five principal strategies of coping with uncertainty: reducing uncertainty (e.g. by collecting additional information); assumption-based reasoning (filling gaps in firm knowledge by making assumptions that go beyond directly available data); weighing pros and cons (of at least two competing alternatives); forestalling (developing an appropriate response or response capabilities to anticipate undesirable contingencies); and suppressing uncertainty (e.g. by ignoring it or by relying on unwarranted rationalization). Similar lists of coping strategies were reported by Allaire and Firsirotu (1989), Janis and Mann (1977), Klein (1998), and Shapira (1995).

Cross-tabulation of the three types of uncertainty with the five strategies of coping revealed that inadequate understanding was principally associated with reduction, lack of information was principally associated with assumption-based reasoning, and conflicted alternatives were principally associated with weighing pros and cons. Forestalling and suppression were equally likely to be used with all three types of uncertainty. Integration of these findings with several models of naturalistic decision making produced the RAWFS heuristic (the acronym designates the five coping strategies), a descriptive model of how decision makers cope with uncertainty.

Although the RAWFS heuristic is descriptive, the logic of its pattern of contingent coping has a certain normative flavor: begin by trying to reduce uncertainty by collecting additional information ('hard facts'), use assumptions to fill gaps in understanding if that's not feasible, compare the merits of competing alternatives if more than one is available, retain a back-up alternative to guard against undesirable contingencies, and resort to suppression only as a last resort.

The Recognition/Metacognition (R/M) model explicates the prescriptive facet that is implicit in any descriptive model of deliberate goal-directed action (Cohen *et al.*, 1996). Similar to the RPD model, the R/M model assumes that naturalistic decision making relies primarily on pattern matching (Cohen *et al.*, 1996). Different from the RPD model, the R/M model focuses on what happens when recognition fails: if stakes are high and time is available, decision makers revert to assumption-based reasoning which, as elaborated in the model, consists of meta-cognitive processes of critical thinking by which decision makers identify and correct gaps in situation awareness and action plans owing to incomplete or conflicting information, inconsistent goals, and unwarranted assumptions.

The R/M model served Cohen and his associates in the development of a generic prescriptive procedure which they labeled STEP (Construct a Story, Test, Evaluate and Plan; Cohen and Freeman, 1997; Cohen *et al.*, 2000). STEP can be applied to improve performance on any decision task that involves perceptual input. For example, based on interviews with active-duty naval officers on their experiences in the Persian Gulf, the Gulf of Sidra, and elsewhere (Kaempf *et al.*, 1996), Cohen and his associates developed a training program for decisions that concern hostile intent in ambiguous situations (i.e. whether or not to engage an approaching air or sea contact whose intent is unknown under conditions of undeclared hostility). This program illustrates how a descriptive model of proficient performance (the R/M model) can be used for prescriptive purposes.

- *Story* Even pattern-matching that yields only vague recognition generates a tentative assessment regarding the nature of the situation, which can be enhanced by construction of a complete Story that recounts past, present, and future events consistent with it. The first component of the Hostile Intent STEP module trains officers in the construction of such stories.
- *Test* Stories are used to test the plausibility of initial assessments by comparison of implications and expectations derived from them with what is known or observed about the situation. When evidence

appears to conflict with an assessment, stories are revised to incorporate all available information into the most complete and plausible account possible. The second component of STEP trains decision makers to spot and correct gaps in stories owing to incomplete evidence and unwarranted underlying assumptions.

- **Evaluate** In the third phase of STEP, decision makers are trained to use a devil's advocate technique in which an infallible 'crystal ball' repeatedly insists that the current assessment is wrong and asks for an explanation. When adjusted stories require too many unwarranted assumptions, decision makers may begin the STEP cycle again with an alternative assessment.
- **Plan** Similar to Forestalling in RAWFS, a back-up best model or plan is available to decision makers using STEP at any moment, qualified by awareness of its strengths and weaknesses. The final component of STEP trains decision makers to plan against the possibility that the current best response is wrong.

Similar to RAWFS, STEP captures tactics that decision makers use to cope with uncertainty, without relying on a normative model. Its prescriptive validity has been tested in five different studies (Cohen *et al.*, 1996; Cohen and Freeman, 1997), which showed statistically significant improvement in the outcomes of the decision-making process due to training, as estimated by agreement of assessments and actions with those of experts in the subject matter.

Uncertainty is intimately linked with error: the greater the uncertainty, the greater the probability of making an error. It is thus not surprising that decision errors attracted the attention of NDM researchers (Klein, 1993; Orasanu *et al.*, 1993). More significantly, the treatment of errors is an important issue that distinguishes NDM from BDT.

THE CONCEPT OF ERROR

Within the framework of BDT, errors are operationally defined as failures to adhere to normative models such as Expected Utility theory and Bayesian statistics. Analytical normative models of optimal choice provide BDT with a basis for detecting errors as well as an engine for conducting research on 'judgmental biases' which produce sub-optimal decisions. By contrast, NDM lacks analytical criteria that serve as signposts for error. The absence of an analytic normative foundation led Doherty (1993) to claim that 'naturalistic decision making is simply silent on what constitutes an error' (p. 380). Doherty has raised three challenges to the NDM community (Lipshitz, 1997b): (1) What constitutes an error? (2) Has NDM made any positive contribution to the understanding of error? (3) Can NDM researchers detect decision errors without the benefit of hindsight?

Rather than denying the reality of errors, field researchers have given careful study to disasters such as Three Mile Island, Bhopal, airline crashes, and the like. The *Vincennes* shoot-down was one of the prime stimuli for the initiation of the NDM movement. For NDM researchers, an error is a useful concept inasmuch as it serves as a flag alerting us to possibilities where performance can be improved. However, under different conditions, it makes more or less sense to talk about errors. And under some situations, talking about error can be misleading. Therefore, in response to question 1, in situations where there are performance standards, and where skilled personnel show consistent use of strategies, we can use these strategies and methods as a basis for comparison and evaluation (while still allowing for the possibility that a departure from the methods used by experts can be an innovation rather than an error). It may also be useful to study the compensations when a person makes a departure from a preferred method. In many field settings, standards do not exist. Here, errors may have to be initially identified through poor outcomes rather than through processes, as it may be more useful to study the factors that influenced the outcome rather than trying to quantify an error rate.

Instead of prescribing reasoning strategies, NDM considers processes such as ineffective attention management and inadequate problem detection, which are likely to result from factors such as workload

and lack of experience. Further, NDM uses the decision processes of experts as yardsticks for sub-standard performance which can be detected without the benefit of hindsight and as goals for emulation (e.g. STEP above). Obviously, there are domains such as stock selection where 'experts' do not perform particularly well. Shanteau (1992) examined the conditions under which expertise leads to superior performance, a necessary condition for adopting experts' behavior as a normative standard in NDM. Thus, years of experience and formal titles are not a guarantee of expertise.

The answer to Doherty's second question, whether NDM has made any positive contribution to the understanding of error, is that the understanding of human error is one of the cornerstones of the NDM framework. Woods and Cook (1999) have described the wide range of cross-disciplinary investigations into human error. While it is beyond the scope of this article to review this body of work, we can at least mention the study of Reason (1990) on latent failures, and that of Rasmussen (1987) on the distinctions between errors made at different levels of cognitive processing. Rasmussen (1997) describes the organizational forces that typically result in a movement toward the boundaries of safe performance. From this perspective, research on errors has been an important opportunity for the field of NDM to study the linkages between different types of causal factors. Instead of tracing bad outcomes to human error as the end of the inquiry, NDM researchers have learned to treat human errors as the beginning of the investigation. They are less likely to attribute the error to faulty reasoning strategies, preferring to use the error as an indicator of poor training or dysfunctional organizational demands, or flawed design of a human-computer interface in order to reduce the likelihood of errors. While BDT generally tries to understand error as the result of faulty decision processes and reliance on fallible heuristics, NDM generally tries to understand error in a broader context, including insufficient experience. In complex settings, there are times when alternative courses of action need to be considered, and times to proceed with the first reasonable option. As people gain experience, and develop richer mental models, they gain the ability to anticipate problems, and to judge when to perform workarounds from the official procedures.

According to Tversky and Kahneman (1974), people are forced to rely on heuristics because of faulty intuitions regarding probabilistic phenomena, in addition to insufficient processing capacity. These heuristics can result in errors, and it may be tempting to explain some types of error in terms of inappropriate use of heuristics. Nevertheless, we should be cautious in attributing errors to the use of heuristics. Klein (1989) showed that the attribution of decision biases in the *Vincennes* shoot-down was *ad hoc*. The same base rate bias would have been used regardless of whether the error was to shoot down a commercial airliner or to fail to shoot down an attacking Iranian fighter. Therefore, there are times when BDT may rely on hindsight just as NDM does in addressing errors in field settings. If we follow BDT and assert that error is the result of faulty decision processes, it becomes important to find ways to reduce or eliminate errors. However, the NDM view (e.g. Lipshitz, 1997b) is that in unstable settings, people may find it adaptive to use errors as a means of learning. A striving for error-free performance may be maladaptive in such settings. The commission of errors *per se* is not necessarily a problem. We need to consider the consequences of errors, not just the reasoning processes.

The structure of the situation may further mitigate the effects of 'faulty' reasoning. Shanteau (1992) described a situation in which physicians exhibited decision biases, but in this natural setting the constraints of practice make the impact of those biases negligible. Therefore, BDT and NDM have made different sorts of contributions to our understanding of error. BDT has worked at the level of micro-cognition to investigate the nature of error; this work entails carefully controlled experiments. NDM researchers have worked at the macro level to understand the ecology of errors; this work entails a concern for applications.

Doherty's third question was whether NDM researchers could detect decision errors without the benefit of hindsight. The reason why BDT is able to define errors without hindsight is that it can define optimal choices, and optimal choice strategies. However, Klein (in press) argues that the concept of optimization is only meaningful in the context of a tightly controlled setting, where the task is for the subject to arrange the information that has been given. Any attempt to broaden the task may render meaningless the calculation

of optimal choice. Allowing subjects to seek additional information creates an infinite regress because the subject has to estimate the costs and benefits of the effort required in information seeking, prior to seeking it, and then must estimate the costs and benefits of estimating those costs and benefits, and so forth. Allowing subjects to consider real consequences requires an exhaustive cataloguing and calibrating of values, looking at long-term goals as well as immediate goals, and constructing simulations of future states marked by considerable ambiguity and uncertainty. BDT researchers clearly recognize these problems. However, in criticizing NDM research, Doherty appears to suggest that BDT can define decision errors in natural settings without hindsight. This is a different matter from setting up controlled studies where errors can be pre-defined. We would place the burden of proof on the decision analysis community, to demonstrate that it has tools for defining decision errors in a broad range of natural settings, without hindsight.

NDM AND TEAMS

Decision making has been traditionally studied at three levels: individual, group, and organizational. Our focus so far has been on the contribution of NDM to theory and research at the first of these levels. We now turn our focus to its contribution at the next level. As teams play critical roles in accomplishing complex, difficult, and often dangerous tasks, NDM researchers focused their attention on answering two questions: (1) What is effective team decision-making (Orasanu and Salas, 1993; Orasanu, 1997)? (2) What turns a team of experts into an expert team (Salas *et al.*, 1997)? These questions were aimed at understanding how decision making evolves and matures in teams comprised of members with distributed knowledge, information, and expertise. However, NDM scientists were conceptually ill prepared to answer these questions. Why? We elaborate below.

The focus of NDM work was on application and not on theory building. While some could argue that first you need to understand and observe how teams make decisions in order to build a team decision-making theory, in fact one needs both. The observations shape the theory and the theory guides the way one studies team decision making in complex environments. NDM researchers have tended to rely on theories and frameworks from other disciplines (e.g. industrial/organizational psychology, social psychology, cognitive psychology, and engineering). This has served as a good point of departure, and new conceptual developments have emerged directly from the NDM paradigm. This includes concepts such as team situation-awareness (Salas *et al.*, 1995), shared problem assessment (Orasanu, 1997), team mind (Klein, 1998), and shared mental models (Cannon-Bowers *et al.*, 1993). These concepts have advanced our understanding of decision making in complex environments.

For example, team situation-awareness (SA) is crucial for effective decision making. In fact, research has demonstrated that obtaining and maintaining SA in teams is far more complex than in individuals. Team SA is achieved, for example, when team members collect and exchange information earlier and plan farther in advance (Orasanu, 1994) and when team members engage in closed-loop communication. Shared mental models are thought to provide team members with a shared understanding of the task, who is responsible for what, and what the information needs and requirements are. This understanding allows team members to anticipate each other's needs without overt strategizing. Research has shown that teams that possess shared mental models exhibit better communication and better planning, and improve their team decision-making performance (Volpe *et al.*, 1996; Stout *et al.*, 1999). While this NDM-based research has generated some rich theoretical notions, there are two additional important contributions of this research.

First, NDM has brought renewed focus to studying teams in context. While this kind of research had been going on for some time (see Hackman, 1990; Foushee, 1984), NDM researchers became more convinced that to understand team decision-making, it had to be studied in its natural environment. This approach, of course, led in turn to a number of conceptual, methodological, and practical problems.

For example, studying teams in context is expensive, labor-intensive, difficult, and frustrating. Results do not come overnight. While these difficulties are not necessarily unique to team research, there are some additional burdens. It takes a team to study teams in context. Tremendous resources and commitment by all involved (sponsors, users, researchers, managers) are required to study teams in context. It takes the conviction, which most NDM scientists and practitioners endorse, that to enhance team decision-making one has to understand the problems teams confront, the environment and situations they encounter, and the nature of their tasks (see Cannon-Bowers and Salas, 1998; Orasanu, 1997; Salas *et al.*, 1997).

We now know what teamwork consists of (McIntyre and Salas, 1995). Teamwork enables effective team decision making. It is the process by which team members seek, exchange, and synchronize information in order to decide on a course of action. McIntyre and Salas (1995) defined teamwork as 'inclusive of the activities that serve to strengthen the quality of functional interactions, relationships, cooperation, communication and coordination of team members' (p. 27). They concluded that teamwork is constituted of a flexible set of behaviors, namely adaptability, shared situational awareness, performance monitoring and feedback, leadership, and closed-loop communication (i.e. the exchange of successful information from one team member to other team members), all of which have been shown to contribute to effective team decision-making (Cannon-Bowers and Salas, 1998).

We also know how to enhance team decision-making performance. That is, we know how to turn a team of experts into an expert team. Many interventions can be used to enhance team performance. For example, Cannon-Bowers *et al.* (1995) explored the efficacy of a variety of instructional strategies (i.e. task simulation, role training, guided practice, lecture, passive demonstration, and role-playing). More specifically, they identified which instructional strategies would be most effective based upon the context, the task, and the team. Recent research has also uncovered which interventions work and which do not (e.g. Cannon-Bowers and Salas, 1998). The US Navy's multimillion-dollar Tactical Decision Making Under Stress (TADMUS) research program afforded researchers the opportunity to examine theories of decision making in depth (see Collyer and Malecki, 1998, for an overview of the program). Briefly, the interventions introduced were aimed at increasing overall skill levels, introducing trainees to stress during training, and targeting skills which were vulnerable to decay. Findings from this series of studies can be used to help guide future efforts in NDM concerning teams. Specifically, many lessons have been learned with respect to conducting large-scale NDM-based team behavioral research (Salas *et al.*, 1998).

We also know a great deal about aircrews, firefighting teams, and medical teams (Zsambok and Klein, 1997). Aircrews in particular have been studied extensively over the past several decades. Variables such as personality types, status differentials, and speech patterns have been examined to determine their effects on decision making (Orasanu and Salas, 1993). These studies have yielded a vast amount of information, which can be drawn upon for the study of other types of teams. For example, the effects of status differential would almost certainly yield similar results in a surgical team where a tenured surgeon was leading a procedure and less experienced surgeons were assisting.

A second contribution of NDM to teams is the current research aimed at designing, developing, and testing better and richer research tools. We know that we need much better methods and tools to capture the complexity of team performance in context. For example, research efforts are in progress to develop cognitive task analysis tools and procedures for teams (Klein, in press; Blickensderfer *et al.*, in press). Also, NDM scientists are working on ways to develop and test knowledge elicitation techniques to evaluate shared cognition in teams (Cooke *et al.*, 1997). Efforts are likewise being made to improve how we capture team performance in context (Cannon-Bowers and Salas, 1997) and how we can study teams in laboratory settings and still have enough confidence to generalize the findings to the field (Bowers *et al.*, 1992; Jentsch and Bowers, 1998; Johnston *et al.*, 1998).

We have discussed the dilemma of rigor versus relevance that confronts NDM researchers who wish to achieve rigor without the artificial context of controlled laboratory experimentation. For example, the key to

performing rigorous experiments on decision making in BDT is the availability of a definition for optimal choice. However, if this concept is not meaningful in natural settings, as we have argued above, the models and methods of BDT may be similarly restricted.

In sum, the NDM paradigm has focused our attention on *real* teams performing *real* tasks in *real* settings. Further, NDM has required research focused on the *process* by which decisions are made and how information between team members is communicated and coordinated. And so some progress has been made in understanding team performance due to the NDM paradigm. The next section addresses, accordingly, two topics: the range of methods used by NDM researchers, along with their rationales, and the question of rigor as applied to these methods.

METHODOLOGY AND RIGOR IN NDM

Understanding decision making in complex natural environments requires methods devoted to illuminating the roles of domain knowledge, perceptual and cognitive processes, and situation, task, and information management strategies. Most research is conducted in the field, drawing on methods from anthropology, ethnography, cognitive science, and discourse analysis. Efforts typically begin with descriptions of the phenomena, without prejudging what is or should be important to study. Descriptive approaches allow the researcher to examine phenomena in their natural contexts rather than leaping to premature attempts to narrow the focus and to test hypotheses. While field methods dominate, other methods may be used, such as simulation and laboratory techniques.

Field studies

Field observations are critical to NDM research because real-world decisions are embedded in and contribute to ongoing tasks. Researchers must understand the environments that demand decisions, the affordances and constraints of those environments, and the kinds of knowledge and skills needed to respond to those demands. Field observations also provide insights into potential sources of difficulty, error, or non-optimal performance, as well as how the larger system supports the decision maker. Methods used for eliciting knowledge from experts (and sometimes novices) include: structured and unstructured interviews (e.g. Cohen *et al.*, 1996; Klein, 1989), retrospective analysis of critical incidents (e.g. Lipshitz and Strauss, 1997), expert drawing of domain maps, think-aloud protocols (e.g. Xiao *et al.*, 1997), and videos of task performance (Omodei *et al.*, 1997). The tasks and materials may be taken from the actual or simulated work environment, may be generated by the analyst or domain expert, and may be designed to be typical or anomalous, easy or challenging, constrained or unconstrained (e.g. in terms of time or information). Real-time field observations (e.g. DiBello, 1997) involve ethnographic techniques. Observers may work *in situ* with practitioners, asking questions such as 'What are you doing? Why? How do you know what to do?' essentially working as 'cognitive apprentices'. One may also conduct field experiments in which a critical feature of the environment is varied in a way that sheds light on how the practitioner thinks about the task (Roth 1997; Roth *et al.*, 1992; Sarter and Woods, 1995).

A key technique of NDM research is *cognitive task analysis* (CTA). (See Gordon and Gill, 1997, for a recent description.) CTA addresses 'the need to capture the knowledge and processing used by experts in performing their jobs' (Gordon and Gill, 1997, p. 131), as well as 'uncovering actual demands confronting practitioners' (Roth, 1997). A type of CTA that focuses specifically on decision making (rather than on an entire complex task) is the Critical Decision Method (confusingly also labeled CDM: Hoffman *et al.*, 1998; Klein *et al.*, 1989). Based on Flanagan's (1954) Critical Incident Technique, this approach provides insights into challenging or unusual decisions. It involves multi-trial retrospection of a specific incident identified by the participant from personal experience. Probe questions are designed to identify important cues, choice

points, options, action plans, and the role of experience. As described by Hoffman *et al.* (1998) three 'sweeps' approach the event from varying perspectives. 'Timeline' verification with decision point identification serves to structure the account into meaningfully ordered segments. Progressive deepening leads to a comprehensive, detailed, and contextually rich account of the incident. 'What-if' queries serve to identify potential errors, alternative decision-action paths, and expert-novice differences' (p. 6.). Products from the CDM analyses include situation assessment records, timelines, and decision requirements.

Full CDM procedures have been used in over 30 studies in domains as diverse as clinical nursing, systems analysis, instructional design, graphic interface design, corporate management, and military planning, command and operations. Products resulting from these analyses include materials that can be used for training, taxonomies of informational or diagnostic cues, and as a basis for assessing skill levels. Decision requirement tables can provide insights into similarities among tasks in terms of their cognitive requirements.

Simulations

Simulated tasks elicit behavior that is similar to what might be seen in an actual situation, but without the risks often present in those environments. Simulations may be extremely high-fidelity, such as aircraft cockpits (e.g. Orasanu and Fischer, 1997), or low-fidelity, such as several process control tasks (Roth *et al.*, 1992) or medical decision tasks (Gaba, in press). Realistic features can be built in, such as temporal parameters, distractions, and workload, and subjects' behavior can be analyzed as a function of relevant factors, such as differences in levels of experience or personality (Cohen and Freeman, 1986; Chidester *et al.*, 1990), or availability of tools or aids (Roth *et al.*, 1987; Woods, 1993).

Laboratory techniques

Salas *et al.* (1995) argued that NDM both can and should be studied in the lab as well as 'in the wild', although doing so means giving up some of the contextual features that define the phenomena in the real world. In fact, NDM researchers have used laboratory methods when understanding of decision making in a particular domain has advanced to a point at which predictions can be made on how decisions are made in meaningful and familiar contexts. For example, Fischer and Orasanu ('Experience and role effects on pilots' interpretations of aviation problems', submitted, 1998) used a sorting task followed by hierarchical clustering and multidimensional scaling to validate aspects of their aviation decision process model, as well as to determine whether the same dimensions were used by captains and first officers to interpret flight decision situations. Klein *et al.* (1995) studied chess players and confirmed a prediction from his recognition-primed decision model, namely that chess masters would generate acceptable moves as the first ones retrieved, in contrast to lower level players, who would engage in more extensive search.

Laboratory experimentation involving large *Ns*, random assignment of subjects to experimental and control conditions, hypothesis testing, and sophisticated statistical tests to evaluate data are permitted in NDM. Still, most questions and types of decisions with which NDM researchers are concerned are not amenable to this type of approach. Consequently, NDM researchers wittingly forgo the type of rigor that guides laboratory studies in order to study decision-making performance in the richness of actual task environments. As Woods (1993) noted, to the degree that decision strategies are task-contingent, one must study the decisions in context. NDM researchers do not yet know enough about task features in most domains to design laboratory studies that will not change the phenomena of interest. Hammond (1988) also emphasized the need to develop a theory of tasks in order to advance our understanding of performance in rich task domains. Considering that laboratory studies are often not suitable for NDM research and that the accepted canon of rigor implicitly assumes laboratory methodology, has NDM given up on the issue of rigor?

Issues of scientific rigor in NDM research

NDM methodology has been criticized as being 'soft' (Yates, in press). This appears to mean that researchers do not adhere to the methods and standards appropriate for laboratory-based experiments. Just as the methods must be suited to the research questions, the criteria for judging the quality of the studies must be appropriate for the methods used. The central question of rigor is whether the methods used to collect and analyze data support the conclusions that are drawn. Researchers working in the field have been just as concerned with issues of data quality and adequacy as those working in the laboratory.

Researchers using cognitive task analysis have asked: How 'good' are the products of a CTA and how can you tell? Concern is expressed over variation in the information generated by different CTA methods (Gordon and Gill, 1997). Are the products biased in any systematic way? How comprehensive, inclusive, and precise are the data?

Hoffman *et al.* (1998) reviewed numerous studies based on the Critical Decision Method (Klein *et al.*, 1989) from the point of view of reliability, validity, and efficiency. To determine reliability they investigated how consistent participants were in reporting the same events, details, or gist of events in a retelling. Retest reliability by fireground commanders across several months averaged 82%. Another reliability check addressed the coding of reported events: Do independent analysts generate similar results from the raw data? Intercoder agreements across a number of similar studies with different participants averaged 85% or better.

NDM researchers are also concerned with the validity of verbal reports: Are distortions introduced in performance of the task due to thinking aloud and limitations on ability to introspect about one's own cognitive processes (cf. Ericsson and Simon, 1984)? Despite these concerns, think-aloud protocols have been used extensively in studies of expert/novice differences (e.g. Chi *et al.*, 1981). Retrospective report techniques raise concern about effects of memory limitations on the data (Loftus, 1996). These concerns suggest that multiple approaches should be included to counteract the limitations of a single method. As Hoffman *et al.* (1998) pointed out, CTA using any method is not like the mining of gold ore; rather, it is knowledge co-discovery or co-creation.

The validity issue also must be addressed from another perspective that deals with the broader question of what NDM research is trying to learn. To the extent that it focuses on *interpretations* and *definitions* of situations by expert decision makers, and the impact of those interpretations on task performance, traditional definitions of validity do not hold:

Reliability, falsifiability, and objectivity are neither trivial nor irrelevant, but they must be understood as particular ways of warranting validity claims rather than as universal, absolute, guarantors of truth. They are rhetorical strategies (Simons, 1989) that fit one model of science, experimental hypothesis testing and so forth. . . . They are literally irrelevant to inquiry-guided research [a generic term denoting research in naturalistic settings that typically uses qualitative methodologies] which does not "test hypotheses", 'measure variation' on quantitative dimensions or "test" the significance of findings with statistical procedures simply because there is nothing in these studies to which to apply them (Mishler, 1990, pp. 419–436).

Using standards of rigor which are suitable for experimentation to evaluate studies that involve observational methods is clearly inappropriate. Just as research methods should be made to fit research questions and not vice versa (Kaplan, 1964), research methods should drive the selection of evaluation criteria and not vice versa.

If the standard criteria are not used to evaluate laboratory studies, then what criteria should be used? Mishler (1990) suggests that inquiry-guided research studies be evaluated according to the criteria of credibility and transferability. *Credibility* refers to the extent to which a study's findings and conclusions are warranted. It is established through information about (a) significance of the research *questions*, (b) methods for data collection and analysis and data upon which answers were based, (c) *suitability* of the methods to the research questions and the research settings, (d) *plausibility* of the answers, and (e)

reasonableness of the assumptions underlying the choice of methods and interpretation of the data. Unfortunately, for researchers who hope to anchor science in a firm foundation of objective knowledge, questions regarding the different facets of credibility, irrespective of the particular methodology employed, are answerable only by judgment calls.

Transferability refers to the extent to which a study's findings and conclusions hold in other settings. It is not based on extrapolation from sample to population based on random sampling and statistical tests, but on a case-to-case translation based on similarity in their significant features (Firestone, 1993). Thus, the notion of transfer requires detailed description of features of the situation, which would be obtained from field studies.

With respect to evaluating rigor, Howe and Eisenhardt (1990, p. 6) point out, 'Failing to follow a given theoretical perspective or methodological convention does not necessarily diminish the warrant of the conclusions drawn'. The central question remains: How good are the data obtained using NDM methods for answering the questions posed by NDM researchers? We might turn the question around and ask: Could traditional laboratory methods do a better job of answering the questions posed by NDM researchers than the methods currently in use? How could their methods be used productively?

As Yates (in press) points out, NDM and traditional decision researchers are looking at different phenomena. They both call it decision making and assume that their own methods apply to the study of the other's problems. This may well be a mistake, if in fact they are talking about apples and oranges. Traditional decision research focuses on theory building and testing, and is concerned with choice and conflict. NDM researchers seek to understand "cognition in the wild" (Hutchins, 1995). We suggest that the scientific rigor and credibility of each must be judged by standards appropriate to each venture.

IN CONCLUSION: CONTRIBUTIONS AND FUTURE CHALLENGES FOR NDM

To take stock of NDM, we reviewed some of the work which has been performed within this framework in the last decade. Our review highlighted the contributions that according to one 'outsider' NDM made to the study of decision making (Yates, in press): the identification of important areas of inquiry hitherto neglected (e.g. complex and dynamic decision processes in naturalistic settings); the introduction of new models (e.g. recognition-primed decisions) and conceptualizations (e.g. of uncertainty and error); the introduction of new methods (e.g. Critical Decision Method); and the recruitment of applied investigators into the field. Above all, as the distinctive characteristics of NDM show, NDM contributed a new perspective on how decisions (broadly defined as committing oneself to a certain course of action) are made. This leads us to believe that NDM is a promising research paradigm to study decision making, linking this field to applied cognition, problem solving, and expertise. At the same time, as Yates (in press) points out, there are also significant challenges ahead.

The first challenge is to develop NDM to be a better science simultaneously focused on solving real-world problems and developing theory built on sound findings, tools, and principles. To this end NDM needs more empirical studies applying appropriately rigorous methodology. Progress in this direction can be achieved via three complementary routes:

- (1) Balance results from field qualitative studies with findings from traditional experimental work (e.g. Klein *et al.*, 1995; Cannon-Bowers and Salas, 1998).
- (2) Develop simulation methods which allow observation of complex decision processes under controlled conditions (e.g. Orasanu *et al.*, 1998; Waag and Bell, 1997).
- (3) Develop better understanding of and methods for rigorous observation (Lipshitz, in press) and knowledge elicitation (Hoffman *et al.*, 1998) of decision making in naturalistic settings.

The availability of more and better empirical research should help NDM meet its second challenge, namely the development of more comprehensive models and theories and well defined boundary conditions for what

NDM is and what it is not (Cannon-Bowers *et al.*, 1996). The ultimate theoretical challenge for NDM, according to these writers, is to specify the 'link between the nature of the task, person, and environment on the one hand and the various psychological processes and strategies involved in naturalistic decisions on the other' (p. 202).

Finally, a third challenge for NDM is to start consolidating its applications. Five years ago we faced the questions of what NDM means, and whether NDM has an impact. We have been busy developing applications since then (Zsambok and Klein, 1997; Salas and Klein, in press). Moreover, we have enjoyed some measure of success. Now we need to converge on some of the more promising types of applications, and conduct careful evaluations to better demonstrate the efficacy of our applications.

In sum, NDM faces challenges which it is well positioned to confront. Its success depends on the viability of NDM's assumptions, theories, methods, empirical work, and applications. This, in turn, should foster a fruitful dialogue among the various three-letter approaches to decision making, thus opening the field to a wider range of issues and a richer set of models, explanations, and applications. In taking stock of NDM we hope to have contributed towards this goal.

ACKNOWLEDGMENTS

We thank Professor Robert Hoffman for his helpful comments.

REFERENCES

- Allaire, Y., & Firsirotu, M.E. (1989). Coping with strategic uncertainty. *Sloan Management Review*, **30**(3), 7–16.
- Argyris, C. (1993). *Knowledge for action*. San Francisco, CA: Jossey-Bass.
- Beach, L.R. (1990). *Image theory: Decision making in personal and organizational contexts*. London: Wiley.
- Beach, L.R. (1993). Broadening the definition of decision making: the role of prechoice screening of options. *Psychological Science*, **4**, 215–220.
- Beach, L.R. (1997). *The psychology of decision making*. London: Sage.
- Beach, L.R., & Mitchell, T.R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, **3**, 439–449.
- Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, **5**, 175–192. (English translation by Sommer, L. (1954). Exposition of a new theory of the measurement of risk. *Econometrica*, **22**, 23–36.)
- Blickensderfer, E.L., Cannon-Bowers, J.A., Salas, E., & Baker, D.P. (in press). Analyzing knowledge requirements in team tasks. In S. Chipman, J.M. Schraagen, & V. Shalin (Eds.). *Cognitive team task analysis*. Mahwah, NJ: Erlbaum.
- Bowers, C.A., Salas, E., Prince, C., & Brannick, M. (1992). Games teams play: a method for investigating team coordination and performance. *Behavior Research Methods, Instruments, and Computers*, **24**, 503–506.
- Calderwood, R., Klein, G.A., & Crandall, B.W. (1988). Time pressure, skill, and move quality in chess. *American Journal of Psychology*, **101**, 481–491.
- Cannon-Bowers, J.A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N.J. Castellan (Ed.). *Individual and group decision making: current issues* (pp. 221–246). Hillsdale, NJ: Erlbaum.
- Cannon-Bowers, J.A., & Salas, E. (1997). A framework for developing team performance measures in training. In M.T. Brannick, E. Salas, & C. Prince (Eds.). *Team performance assessment and measurement: theory, methods, and applications* (pp. 45–62). Mahwah, NJ: Erlbaum.
- Cannon-Bowers, J.A., & Salas, E. (1998). Individual and team decision making under stress: theoretical underpinnings. In J.A. Cannon-Bowers, & E. Salas (Eds.). *Making decisions under stress: implications for individual and team training* (pp. 17–38). Washington, DC: APA Press.
- Cannon-Bowers, J.A., & Salas, E. (Eds.). (1998). *Decision making under stress: implications for training and simulation*. New York: American Psychological Association.
- Cannon-Bowers, J.A., Salas, E., & Pruitt, J.S. (1996). Establishing the boundaries of a paradigm for decision research. *Human Factors*, **38**, 193–205.

- Carroll, J.S. (1980). Analyzing decision behavior: the magician's audience. In T. Wallsten (Ed.). *Cognitive processes in choice and decision behavior* (pp. 68–76). Hillsdale, NJ: Erlbaum.
- Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, **5**, 121–152.
- Chidester, T.R., Kanki, B.G., Foushee, H.C., Dickinson, C.L., & Bowles, S.V. (1990). *Personality factors in flight operations: Volume I. Leader characteristics and crew performance in a full-mission air transport simulation*. Technical Memorandum NASA TM-102259. Moffett Field, CA: Ames Research Center, April.
- Cohen, M.S., & Freeman, J.T. (1997). Understanding and enhancing critical thinking in recognition-based decision making. In R. Flin, & L. Martin (Eds.). *Decision making under stress: emerging themes and applications* (pp. 161–169). Aldershot: Ashgate.
- Cohen, M.S., Freeman, J.T., & Thompson, B. (1998). Critical thinking skills in tactical decision making: a model and a training strategy. In J.A. Cannon-Bowers, & E. Salas (Eds.). *Decision making under stress: implications for training and simulation*. Washington, DC: American Psychological Association.
- Cohen, M.S., Freeman, J.T., & Wolf, S. (1996). Meta-recognition in time stressed decision making: recognizing, critiquing and correcting. *Human Factors*, **38**, 206–219.
- Cohen, M.S., Thompson, B.B., Adelman, L., Bresnick, T.A., Shastri, L., & Riedel, A. (2000). *Training critical thinking for the battlefield. Volume II: Training system and evaluation*. Arlington, VA: Cognitive Technologies, Inc.
- Collyer, S.C., & Melecki, G.S. (1998). Tactical decision making under stress: History and overview. In J.A. Cannon-Bowers & E. Salas (Eds.). *Making decision under stress: Implications for individual and team training*. Washington, DC: American Psychology Association.
- Cook, R.I., & Woods, D.D. (1994). Operating at the sharp end: the complexity of human error. In M.S. Bogner (Ed.). *Human error in medicine*. Hillsdale, NJ: Erlbaum.
- Cooke, N.J., Stout, R.J., & Salas, E. (1997). Broadening the measurement of situation awareness through cognitive engineering methods. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*. Santa Monica, CA, 215–219.
- Coombs, C.H., Dawes, R.M., & Tversky, A. (1971). *Mathematical psychology: an elementary introduction*. Englewood Cliffs, NJ: Prentice Hall.
- Crandall, B., & Getchell-Reiter, K. (1993). Critical decision method: a technique for eliciting concrete assessment indicators from the intuition of NICU nurses. *Advances in Nursing Science*, **16**(1), 42–51.
- Cyert, R., & March, J. (1963). *A behavioral theory of the firm*. Englewood Cliffs, NJ: Prentice Hall.
- Dawes, R.M. (1988). *Rational choice in an uncertain world*. New York: Harcourt Brace Jovanovich.
- De Groot, A.D. (1965). *Thought and choice in chess*. The Hague: Mouton.
- Dewey, J. (1933). *How we think*. Boston, MA: D.C. Heath.
- Dibello, L. (1997). Exploring the relationship between activity and expertise: paradigm shifts and decision defaults among workers learning material requirements planning. In C.E. Zsambok & G.A. Klein (Eds.). *Naturalistic decision making* (pp. 163–174). Mahwah, NJ: Erlbaum.
- Doherty, M.E. (1993). A laboratory scientist's view of naturalistic decision making. In G.A. Klein, J. Orasanu, R. Calderwood & C. Zsambok (Eds.). *Decision making in action: models and methods* (pp. 362–388). Norwood, CT: Ablex.
- Edwards, E. (1954). The theory of decision making. *Psychological Bulletin*, **51**, 380–417.
- Einhorn, H.J., & Hogarth, R.M. (1986). Decision making under ambiguity. *Journal of Business*, **59**, 225–250.
- Endsley, M.R. (1997). The role of situation awareness in naturalistic human decision making. In C. Zsambok, & G.A. Klein (Eds.). *Naturalistic decision making*. Hillsdale, NJ: Erlbaum.
- Ericsson, K.A., & Leman, A.C. (1996). Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, **47**, 273–305.
- Ericsson, K.A., & Simon, H.A. (1984). *Protocol analysis*. Cambridge, MA: MIT Press.
- Firestone, W.A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher*, **22**, 16–23.
- Fischhoff, B. (1982). Debiasing. In D. Kahneman, P.A. Slovic, & A. Tversky (Eds.). *Judgment under uncertainty: heuristics and biases* (pp. 422–444). New York: Cambridge University Press.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, **51**, 327–358.
- Flin, R. (1996). *Sitting in the hot seat: leaders and teams for critical incident management*. Chichester: Wiley.
- Flin, R., Salas, E., Strub, M., & Martin, L. (Eds.). (1997). *Decision making under stress: emerging themes and applications*. Aldershot: Ashgate.
- Foushee, H.C. (1984). Dyads and triads at 35,000 feet: factors affecting group process and aircrew performance. *American Psychologist*, **39**, 885–893.

- Funder, D.C. (1987). Errors and mistakes: evaluating the accuracy of social judgment. *Psychological Bulletin*, **101**, 75–90.
- Gaba, D. (in press). Applying crew resource management training to team decision making of medical personnel. In E. Salas, & G.A. Klein (Eds.). *Research, methods and applications of naturalistic decision making principles*. Mahwah NJ: Erlbaum.
- Gigerenzer, G., & Todd, P.M. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.
- Gordon, S.E., & Gill, R.T. (1997). Cognitive task analysis. In C.E. Zsambok, & G.A. Klein (Eds.). *Naturalistic decision making* (pp. 131–140). Mahwah, NJ: Erlbaum.
- Grandori, A. (1984). A prescriptive contingency view of organizational decision-making. *Administrative Science Quarterly*, **29**, 192–209.
- Hackman, J.R. (Ed.). (1990). *Groups that work (and those that don't): creating conditions for effective teamwork*. San Francisco, CA: Jossey-Bass.
- Hammond, K.R. (1988). Judgment and decision making in dynamic tasks. *Information and Decision Technologies*, **14**, 3–14.
- Hammond, K.R. (1993). Naturalistic decision making from a brunswikian viewpoint: past, present, future. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.). *Decision making in action: models and methods* (pp. 205–227). Norwood, CT: Ablex.
- Hoffman, R., Shadbolt, N.R., Burton, A.M., & Klein, G.A. (1995). Eliciting knowledge from experts: a methodological analysis. *Organizational Behavior and Human Decision Processes*, **62**, 129–158.
- Hoffman, R.R., Crandell, B., & Shadbolt, N. (1998). Use of critical decision method to elicit expert knowledge: A case study in the methodology of expert task analysis. *Human Factors*, **40**, 254–276.
- Hogarth, R.M. (1987). *Judgment and choice*. London: Wiley.
- Howe, K., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research: a prolegomenon. *Educational Researcher*, **(5)**, 1–11.
- Humphreys, P., & Berkeley, D. (1985). Handling uncertainty: levels of analysis of decision problems. In G. Wright (Ed.). *Behavioral decision making* (pp. 257–282). New York: Plenum Press.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Janis, I.L., & Mann, L. (1977). *Decision making: a psychological analysis of conflict, choice and commitment*. New York: Free Press.
- Jentsch, F.G., & Bowers, C.A. (1998). Evidence for the validity of PC-based simulations in studying aircrew coordination. *The International Journal of Aviation Psychology*, **8**, 195–318.
- Johnston, J.A., Poirier, J., & Smith-Jentsch, K.A. (1998). Decision making under stress: creating a research methodology. In J.A. Cannon-Bowers, & A. Salas (Eds.). *Making decisions under stress: implications for individual and team training* (pp. 39–59). Washington, DC: APA Press.
- Kaempf, G.F., Klein, G., Thordesen, M.L., & Wolf, S. (1996). Decision making in complex command-and-control environments. *Human Factors*, **38**, 206–219.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: heuristics and biases*. New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, **47**, 263–291.
- Kaplan, A. (1964). *The conduct of inquiry*. Scranton, PA: Chandler.
- Klein, G.A. (1989). Do decision biases explain too much? *Human Factors Society Bulletin*, **22**, 1–3.
- Klein, G.A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.). *Decision making in action: models and methods*. Norwood, CT: Ablex.
- Klein, G.A. (1998). *Sources of power: how people make decisions*. Cambridge, MA: MIT Press.
- Klein, G.A. (in press). Cognitive team task analysis. In S. Chipman, J.M. Schraagen, & V. Shalin (Eds.). *Cognitive team task analysis*. Mahwah, NJ: Erlbaum.
- Klein, G.A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, **19**, 462–472.
- Klein, G.A., & Crandall, B.W. (1995). The role of mental simulation in naturalistic decision making. In P. Hancock, J. Flach, J. Caird, & K. Vincente (Eds.). *Local applications of the ecological approach to human-machine systems* (Vol. 2) (pp. 324–358). Hillsdale, NJ: Erlbaum.
- Klein, G.A., Orasanu, J., Calderwood, R., & Zsambok, C.E. (Eds.). (1993). *Decision making in action: models and methods*. Norwood, CT: Ablex.
- Klein, G.A., Wolf, S., Militello, L., & Zsambok, C.E. (1995). Characteristics of skilled option generation in chess. *Organization Behavior and Human Decision Processes*, **62**, 63–69.

- Larkin, J.H., McDermott, J., Simon, H.A., & Simon, D.P. (1980). Expert and novice performance in solving physics problems. *Science*, **208**, 1335–1342.
- Lipshitz, R. (1993). Converging themes in the study of decision making in realistic settings. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.). *Decision making in action: models and methods* (pp. 103–137). Norwood, CT: Ablex.
- Lipshitz, R. (1994). Decision making in three modes. *Journal for the Theory of Social Behavior*, **24**, 47–66.
- Lipshitz, R. (1997a). Coping with uncertainty: beyond the reduce, quantify and plug heuristic. In R. Flin, E. Sala, M. Strub, & L. Martin (Eds.). *Decision making under stress: emerging themes and applications* (pp. 149–160). Aldershot: Ashgate.
- Lipshitz, R. (1997b). Naturalistic decision making perspectives on decision errors. In C.E. Zsambok, & G.A. Klein (Eds.). *Naturalistic decision making* (pp. 151–162). Mahwah, NJ: Erlbaum.
- Lipshitz, R. (in press). Puzzle seeking and model building on the fire ground. In E. Salas, & G.A. Klein (Eds.). *Research, methods and applications of naturalistic decision making principles*. Mahwah NJ: Erlbaum.
- Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: a naturalistic decision making analysis. *Organizational Behavior and Human Decision Processing*, **69**, 149–163.
- March, J.G. (1982). Theories of choice and the making of decisions. *Society*, **20**, 29–39.
- March, J.G., & Simon, H.A. (1958). *Organizations*. New York: Wiley.
- McIntyre, R.M., & Salas, E. (1995). Measuring and managing for team performance: emerging principles from complex environments. In R. Guzzo, & E. Salas (Eds.). *Team effectiveness and decision making in organizations* (pp. 149–203). San Francisco: Jossey-Bass.
- Meehl, P.E. (1954). *Clinical vs. Statistical Predictions: Theoretical Analysis and Review of the Evidence*. Minneapolis: University of Minnesota Press.
- Mischler, E.G. (1990). Validation in inquiry-guided research: the role of exemplars in narrative studies. *Harvard Educational Review*, **60**(4), 415–441.
- Montgomery, H. (1988). From cognition to action: the search for dominance in decision making. In H. Montgomery, & O. Svenson (Eds.). *Process and Structure in Human Decision Making*. Wiley: New York; 471–483.
- Newell, A., & Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.
- Omodei, M., Wearing, A., & McLennan, J. (1997). Head-mounted video recording: a methodology for studying naturalistic decision making. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.). *Decision making under stress: emerging themes and applications* (pp. 161–169). Aldershot: Ashgate.
- Orasanu, J. (1994). Shared problem models and flight crew performance. In N. Johnston, N. McDonald, & R. Fuller (Eds.). *Aviation psychology in practice* (pp. 255–285). Aldershot: Ashgate.
- Orasanu, J. (1997). Stress and naturalistic decision making: strengthening the weak links. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.). *Decision making under stress: emerging themes and applications* (pp. 49–160). Aldershot: Ashgate.
- Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.). *Decision Making in Action: Models and Methods* (pp. 3–20). Norwood, CT: Ablex.
- Orasanu, J., Dismukes, R.K., & Fischer, U. (1993). Decision errors in the cockpit. In J. Smith (Ed.). *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, **1** (pp. 363–367). Santa Monica, CA: Human Factors and Ergonomics Society.
- Orasanu, J., & Fischer, U. (1997). Finding decisions in natural environments. In C.E. Zsambok, & G.A. Klein (Eds.). *Naturalistic decision making* (pp. 434–458). Hillsdale, NJ: Erlbaum.
- Orasanu, J., Fischer, U., McDonnell, L.K., Davison, J., Haars, K.E., Villeda, E., & Vanaken, C. (1998). How do flight crews detect and prevent errors? Findings from a flight simulation study. *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 191–195). Santa Monica, CA: HFES.
- Orasanu, J., & Salas, E. (1993). Team decision making in complex environments. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.). *Decision making in action: models and methods* (pp. 327–345). Norwood, NJ: Ablex.
- Patel, V.L., & Groen, G.J. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science*, **10**, 91–116.
- Payne, J.W., Johnson, E.J., Bettman, R., & Coupland, E. (1990). Understanding contingent choice: a computer simulation approach. *IEEE Transactions on Systems, Man and Cybernetics*, **20**, 296–309.
- Pennington, N., & Hastie, R. (1993). A theory of explanation-based decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.). *Decision making in action: models and methods* (pp. 188–201). Norwood, CT: Ablex.

- Pruitt, J.S., Cannon-Bowers, J.A., & Salas, E. (1997). In search of naturalistic decisions. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: Emerging themes and applications* (pp.29–42). Aldershot, UK: Ashgate.
- Rasmussen, J. (1987). The definition of human error and a taxonomy for technical system design. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New Technology and Human Error*. New York: Wiley.
- Rasmussen, J. (1997). Merging paradigms: decision making, management, and cognitive control. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.). *Decision making under stress: emerging themes and applications* (pp. 67–84). Aldershot: Ashgate.
- Reason, J. (1990). *Human Error*. Cambridge: Cambridge University Press.
- Roth, G. (1997). From individual and team learning to systems learning. In S. Cavalieri & D. Fearn (Eds.), *Managing in organizations that learn*. Cambridge, MA: Blackwell.
- Roth, E.M., Woods, D.D. & Pople, H.E. (1992). Cognitive Simulation as a tool for cognitive tasks analysis. *Ergonomics*, **35**, 1163–1198.
- Russo, E.J., & Schoemaker, P.J.H. (1987). *Decision traps: ten barriers to brilliant decision making and how to overcome them*. New York: Doubleday.
- Salas, E., Cannon-Bowers, J.A., & Johnston, J.H. (1997). How can you turn a team of experts into an expert team? Emerging training strategies. In C.E. Zsambok, & G.A. Klein (Eds.). *Naturalistic decision making* (pp. 359–370). Mahwah, NJ: Erlbaum.
- Salas, E., & Klein G.A. (in press). *Research, methods and applications of naturalistic decision making principles*. Mahwah, NJ: Erlbaum.
- Salas, E., Prince, C., Baker, D.P., & Shrestha, L. (1995). Situation awareness in team performance: implications for measurement and training. *Human Factors*, **37**, 123–136.
- Sarter, N.B., & Woods, D.D. (1995). How in the world did we get into that mode? Mode error and awareness in supervisory control. *Human Factors*, **37**, 5–19.
- Savage, L.J. (1954). *The foundations of statistics*. New York: Wiley.
- Searle, J.R. (1995). The mystery of consciousness. *The New York Review of Books*, 2 November, 60–66.
- Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, **53**, 252–266.
- Shapira, Z. (1995). *Risk taking: a managerial perspective*. New York: Russell Sage.
- Simon, H.A. (1957). *Administrative behavior*. New York: Free Press.
- Simon, H.A. (1978). Rationality as process and as product of thought. *American Economic Association*, **68**, 1–16.
- Simons, H.W. (1989). (Ed.). *Rhetoric in the Human Sciences*. Newbury Park, CA: Sage.
- Smith, G.F. (1997). Managerial problem solving: a problem-centered approach. In G.A. Klein, & C.E. Zsambok (Eds.). *Naturalistic Decision Making* (pp. 371–382). Mahwah, NJ: Erlbaum.
- Stout, R.J., Cannon-Bowers, J.A., & Salas, E. (in press). Team situational awareness (SA): Cue-recognition training. In M. McNeese, M.R. Endsley, & E. Salas (Eds.), *New Trends in cooperative activities*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Teigen, K.H. (1996). Decision making in two worlds. *Organizational Behavior and Human Decision Processes*, **65**, 249–251.
- Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, **79**, 281–299.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, **185**, 1124–1131.
- Volpe, C.E., Cannon-Bowers, J.A., Salas, E., & Spector, P. (1996). The impact of cross training on team functioning. *Human Factors*, **38**, 87–100.
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. New York: Wiley.
- Waag, W.L., & Bell, H.H. (1997). Situation assessment and decision making in skilled fighter pilots. In C.E. Zsambok, & G.A. Klein (Eds.). *Naturalistic decision making* (pp. 247–256). Mahwah, NJ: Erlbaum.
- Wagenaar, W.A., Keren, G., & Lichtenstein, S. (1988). Islanders and hostages: deep and surface structures of decision problems. *Acta Psychologica*, **67**, 175–188.
- Woods, D.D. (1993). Process-tracing methods for the study of cognition outside of the experimental psychology laboratory. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.). *Decision making in action: models and methods* (pp. 228–251). Norwood, NJ: Ablex.
- Woods, D.D., & Cook, R.I. (1999). Perspectives on human error: hindsight biases and local rationality. 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*. New York: Wiley.
- Xiao, Y., Milgram, P., & Doyle, D.J. (1997). Capturing and modeling planning expertise in anesthesiology: results of a field study. In G.A. Klein, & C.E. Zsambok (Eds.). *Naturalistic Decision Making* (pp. 197–205). Mahwah, NJ: Erlbaum.

- Yates, J.F. (in press). 'Outsider' impressions of naturalistic decision making. In E. Salas, & G.A. Klein (Eds.). *Research methods and applications of naturalistic decision making principles*. Mahwah, NJ: Erlbaum.
- Zsambok, C.E. (1997). Naturalistic decision making: where are we now? In C.E. Zsambok, & G.A. Klein (Eds.). *Naturalistic decision making* (pp. 3–16). Mahwah, NJ: Erlbaum.
- Zsambok, C.E., & Klein, G.A. (Eds.). (1997). *Naturalistic decision making*. Mahwah, NJ: Erlbaum.

Authors' biographies:

Raanan Lipshitz is an Associate Professor in the Department of Psychology of Haifa University, Haifa, Israel. His research interests include naturalistic decision making (with specific interests in coping with uncertainty and knowledge-driven decision processes), organizational learning, and qualitative methodology.

Gary Klein is Chief Scientist of Klein Associates, Inc., a company he founded in 1978 to improve decision making in individuals and teams. He received his PhD in experimental psychology from the University of Pittsburgh in 1969, taught at Oakland University (1970–1974), and worked as a research psychologist for the US Air Force (1974–1978).

Judith Orasanu, Ph.D., is a Principal Investigator at NASA-Ames Research Center at Moffett Field, CA. For the past decade she has conducted research on naturalistic decision making, focusing on aviation flight crews. Her work has examined crew decision-making processes, crew communication strategies, and the role of cognitive and contextual factors in decision errors associated with aviation accidents.

Eduardo Salas is a Professor of Psychology at the University of Central Florida. He has co-authored over 170 journal articles and book chapters and has co-edited ten books. He is on the editorial boards of *Human Factors*, *Personnel Psychology*, *Military Psychology*, *Group Dynamics*, and the *Journal of Organizational Behavior*. He is a Fellow of the American Psychological Association and of the Human Factors and Ergonomics Society.

Authors' addresses:

Raanan Lipshitz, Department of Psychology, University of Haifa, Haifa, Israel 31905.

Gary Klein, Klein Associated Inc., USA.

Judith Orasanu, NASA Ames, USA.

Eduardo Salas, University of Central Florida, USA.