New York Group Studies Managers' "Mental Models"

Thomas Stewart
Center for Policy Research
SUNY—Albany

I've thrown in with a couple of system dynamicists, and we are going to study how managers' "mental models" are related to their performance on a dynamic task. A regression model of judgment is a kind of mental model, but the term is used in different ways in human factors studies, AI, decision theory (e.g., influence diagrams), and system dynamics. Our concept of mental models draws heavily on system dynamics and, of course, the lens model.

For our study, the task will be managing a social program (Jobs Opportunity and Basic Skills—the JOBS program, authorized in 1988). A lot is riding on this program, and it is not going well. Analysis of the JOBS program suggests that if managers at different levels act reasonably (from their own perspective) the program will fail.

We are going to get groups of students and real managers to role-play JOBS managers. We have a computer model that realistically simulates the JOBS program, and we can measure group and individual success. We will also assess certain features of participants' mental models of the task. We want to test the hypothesis that critical components of mental models (e.g., presence of certain feedback loops and knowledge of delays) will be related to performance. Other dimensions of interest include operator vs. design logic, changes in mental models resulting from experience or training, alternative procedures for assessing mental models, and validity of representations of mental models.

Hammond Preparing Book Manuscript on Uncertain Judgments, Inevitable Error

Kenneth Hammond
Center for Research on Judgment & Policy
University of Colorado, Boulder

For most of the past year I have been working on a book manuscript, tentatively entitled, *Uncertain Judgments, Inevitable Error: Consequences for Policy Formation and Stressful Decisions.*

The vast majority of all human judgments are now, and in the foreseeable future will be, made under conditions of irreducible uncertainty. No one has yet addressed the consequences of these circumstances for the formulation of social policy and/or for stressful decision making. This book provides a brief, innovative, and historical description of our efforts to cope with irreducible uncertainty in the social and ecological circumstances surrounding us, as well as describing current academic efforts to understand human judgment and decision making. In addition, it offers a comprehensive theory that builds on and extends previous work. In Part II the comprehensive theory is applied to social policy formulation as it is undertaken now and as it might be in view of research-based knowledge. The theory is also applied to stressful decision making, a set of circumstances assumed to be prevalent in the foreseeable future.

This book is not a textbook in J/DM, human factors psychology, cognitive psychology, or management science, indeed, not a textbook in any field. It does, however, address problems related to all these disciplines (and possibly others) by virtue of the fact that it (a) offers an unusual history of uncertainty in human judgment and its role in human affairs, (b) presents a comprehensive theory of human judgment, and (c) demonstrates its application to policy formation and stressful decisions.

I am roughly half finished; I hope to finish a reviewable draft by December ($p = .5$).

Together with Reid Hastie and Lewis Harvey I published an article in *Psychological Science* (1992, *3*, 80-87) entitled, "Making Better Use of Scientific Knowledge: Separating Truth from Justice." If Brunswikians will read Leslie Roberts's article, "Science in Court: A Culture Clash" (*Science*, 1992, *257*, 732-736) and Hammond and Adelman's "Science, Values, and Human Judgment" (*Science*, 1976, *194*, 389-396), they will have a new appreciation of the need for research in human judgment.

Brunswikian Research Down Under

Ray Cooksey
Department of Behavioural Studies
University of New England
Armidale, Australia

Two student projects involving the use of SJT methods have been completed in 1992 and one is in the final programming stages. The first project concerned the capturing of judgment policies in staff and executive board members of a local retirement village. The context focused on judgments of the desirability of moving elderly people residing in the village to a full-care nursing home. Relevant cues (such as degree of incontinence, mobility, and self-care ability, were identified through staff interviews; hypothetical profiles were generated and culled for unrepresentative cue configurations. Policies were captured for both residential care staff and executives charged with actually making the decisions regarding when it was time to move a resident to a nursing home. Care staff and executive staff were found to significantly disagree in policy weights and function forms leading to fruitful discussion of how such decisions were actually reached.

The second project was undertaken in Queensland and involved a study of the decisions made by senior officers and residential care workers with respect to disabled people in the Queensland community health program. The crucial decisions concerned the number of hours per day particular clients needed to be supervised and where the client should best be housed (in a residential care facility or a community-based care facility). In this study, real client profiles were available with cues such as ability to handle finances, respecting rights of others, and self-care ability, being quantitatively encoded for judgment. Each client had criterion decisions (number of hours and care facility) on file to which judgments could be compared. As the cue set was highly intercorrelated, policy weights were captured as squared semipartial correlations to handle the multicollinearity problem. Relative weights were computed as the proportion of uniquely explained variation that each cue accounted for. No significant differences were found between residential care worker policies and senior officer policies in the system in terms of judging number of hours of supervision. Both achievement (mean = .67) and matching (mean = .92) were high as was consistency (mean = .89). However, judgments of which care facility would be appropriate were more inconsistent (mean = .78), showed less achievement (mean = .26) and less matching (mean = .61) indicating that these judgments were far harder to make and generated more disagreement among judges. The interesting thing about this project was the resistance and uncertainty expressed by the staff when having their "judgments" analyzed. However, they were reassured that the task was not evaluative in nature and that they would receive feedback regarding their judgment policies.

The final study in preparation concerns the software interlinking of POLICY PC and MOUSELAB computer programs to facilitate the use of MOUSELAB (Johnson, Payne, Schkade, & Bettman, 1989) to gather profile judgments under time pressure conditions. James Hogge's CUER correlated cue generation program is also being used to generate representative judgment profiles. Goals of the study include (a) the validation of policy weights for cues against the time cue information is examined when making a judgment and (b) the effects of time pressure on cue utilization and study time. It is anticipated that policy weights (in a regression sense) will more closely match study time allocation as time pressure increases on the judgment process. Pilot data is currently being collected to test the software bridge between programs.

Decision Making at the End of the Universe

Alex Wearing, Department of Psychology
University of Melbourne

Melbourne has now become a center for psychological decision research in Australia (although the level of competition is not quite Olympian) with the arrival of Leon Mann and his group from Flinders University in Adelaide, South Australia, who are based at the Graduate School of Management.

The Department of Psychology itself has moved into the Faculty of Medicine, and already projects with a medical flavor are either in contemplation or underway, for example, anesthetists' decisions (Jeanette Lawrence), nurses' models of their task environment (Jeanette Lawrence and Alex Wearing), and the decision to screen or not to screen against cervical cancer (Vicki White and Alex Wearing).

Other projects include decision making by expert and novice accountants (Chris Ball, Leon Mann, Keith Houghton, and Alex Wearing), and dynamic decision making, mainly using FIRECHIEF, a simulation of the task facing a fire control office (Mary Omodei and Alex Wearing).

FIRECHIEF simulates any process which is fire-like in its behavior. This program was going to be ready for distribution two years ago, so the present planned completion date of this coming October may be also a vain boast. This time, perhaps, we might make it, as the program is running sweetly on most IBM clones, and a 200+ page manual is almost finished.

Substantively, we are continuing to run studies examining the relationship between various affects and cognition, and their influence on task performance. We have also just started looking at the decision to shopsteal (Jeanette Lawrence and Alex Wearing). There is also a small project jogging along on economic decision making (Maureen Williamson, Shane Thomas, and Alex Wearing). Discussing the economy and its problems is a popular pastime hereabouts.

We have had visitors: Bernd Rohrmann, known to SPUDM goers, has been here for half a year. Previously Josh Klayman spent an extended period here.

Medical Decision Making

C. R. B. Joyce
and
Hans-Ulrich Fisch
Department of Psychiatry
University of Bern

Dick Joyce, whose appointment as Visiting Professor to the University Psychiatric Poliklinik in Bern has another year to run, is now also a Visiting Professor at the Department of Psychology, Royal College of Surgeons in Ireland. In Switzerland he continues to collaborate with the group developing the SEIQoL (Schedule for Evaluation of Individual Quality of Life), which has been described in the journals *Lancet* and *Psychological Medicine*. He also consults for the Swiss National Research Foundation on Alternative Medicines.

Hans-Ulrich Fisch (Department of Psychiatry, University of Bern) and Dick Joyce are involved in an ongoing study on the quality of life in patients with alcoholism. Twenty alcoholics were treated for half a year with disulfiram under supervision by a trustee. Liver function and social adaptation improved during treatment. Quality of life was assessed before the study, three months, and six months after initiation of treatment using the SEIQoL. Quality of life of matched volunteers was also assessed as a control.

• • • • • • • • • •

Marilyn Rothert
College of Nursing
Michigan State University

Our research team is continuing with our NIH grant to develop and test an educational intervention to assist and empower women in the decisions regarding menopause and hormone replacement therapy. We are using cases from our previous judgment study as part of the group intervention. We are measuring satisfaction with and adherence to the decision as well as knowledge, perceptions, and several other variables measured over a 12-month period postintervention. We have completed pilot testing of the instruments and intervention, and in the next few months will be doing the study with 300 women randomly assigned to one of three groups.

We also have some supplemental grants to begin to explore the decision making of low income African American women on this issue. We are beginning to identify the variables for replicating the judgment study with this population.

• • • • • • • • • •

Roy Poses
Medical College of Virginia
Virginia Commonwealth University

The study I am doing with Donna A. Forti and Wally R. Smith of physicians' judgments and decisions for patients with heart failure continues. We will finish enrolling patients at the end of this year, and hope to begin analysis soon after. We described this study in more detail in last year's newsletter.

Mark Chaput de Saintonge, Tom Tape, Bob Wigton, and I designed a rather ambitious study to develop written vignettes to explore how physicians decide to use empiric antibiotics for common infectious problems (sore throats and urinary tract infections); validate these vignettes as predictors of physicians' decisions for actual patients; and then use the vignettes to explore the causes of international practice variation. We submitted our proposal in response to an RFA from the AHCPR for projects on outpatient drug therapy of common problems. We have no idea of our chances this time, but are enthusiastic about pursuing this project one way or another. We believe the time has arrived to apply theories from judgment/decision psychology, and from the Brunswikians in particular to problems in health services research.

Judging Rapport Topic of Study

John Gillis
Department of Psychology
Oregon State University

Frank Bernieri and I are continuing our work on judging rapport, having subjects try to assess the extent to which individuals like or dislike each other from videotapes of their interactions. We have continued to examine the enormous body of data collected in Oregon late last year, and I repeated that study in Greece this spring. We have found out that, (1) while the task is difficult ($R=.70$), rapport can indeed be accurately assessed from even very brief samples of interactions if cues are correctly weighted; but that, (2) as might be expected, subjects, if not given feedback, do not do particularly well, at least if the achievement correlation is the index of performance. (3) The hierarchy of subjects' use of cues is, however, quite similar to the hierarchy of ecological validities. Achievement is diminished because of inconsistent policies and the tendency to weigh heavily a couple more obvious cues—smiling and gesturing—which have in fact, low validity. (4) While there are marked individual differences among subjects in their success at this complex social judgment task, we have been able to identify no pattern of personality variables which predicts their performance. We have spent a lot of time at this last—it has all sorts of implications for clinical work—and have found individual personality indices which correlate with achievement but again no really plausible pattern of such variables.

I've also taken a close look at an earlier study on stress and judgment with a couple of interesting results and recently completed a project which involved Dick Joyce's "Quality of Life" measure developed from SJT. I look forward to saying a few words about all of this at our next meeting.

Judgments in Performance Appraisal Investigated

James Holzworth
Department of Psychology
University of Connecticut, Storrs

At the University of Connecticut (Storrs), my colleague Janet Barnes-Farrell and I are investigating judgment issues in performance appraisal. Janet and I attended JAWS II in Albany last summer, so she now has a better appreciation of the Brunswikian perspective. We have a judgment study under way concerning reactions to disciplinary problems in the workplace and remedies judged to be appropriate. Nature of the rule violation/infraction, worker age, tenure, and some other cues are being varied by case.

Janet and I are also fiddling with some videotapes of waiters and waitresses performing their jobs. We have detailed scripts of the videotapes, and we are interested in how we might perform some kind of test of cognitive continuum theory. We suspect that display format might influence judgments of performance. Perhaps scripts or outlines of work activities will invite more analytical cognition, and videos will invite more intuitive cognition? This research is in the design stage now.

A graduate student, Steve Queen, and I are investigating how computer displays (alphanumeric vs. graphic) of cue values influence multi-attribute evaluations. Data have been collected, but not analyzed. I'll have something to report in November.

Articles Focus on Interface Design

Kim Vicente
Department of Industrial Engineering
University of Toronto

During this past year, I have concentrated on writing several journal articles that were derived from my dissertation, *Supporting Knowledge-Based Behavior Through Ecological Interface Design*. This work was directed at both applied and basic research issues. The primary applied concern was to evaluate a novel theoretical framework for interface design for complex-human machine systems. The basic concern was to better understand the relationship between expertise and memory recall performance. I plan on building on this work by investigating how human adaptation can be influenced by interface design. The idea is to see how various design manipulations can affect the strategies that people develop with practice. This work will be conducted at the Industrial Engineering Department at the University of Toronto, where I moved in July.

CUER Operational

James Hogge
Peabody College
Vanderbilt University

CUER, a computer program to generate cues approximating specified intercorrelations, is now operational. The program is written in GW-BASIC and produces output compatible with POLICY PC. Please contact me (HOGGEJH@vuctrvax or 615-322-8141 or Box 512, Peabody College of Vanderbilt University, Nashville, TN 37203) if you would like a copy.

John Murrell (Homerton College, The University of Cambridge) and I are planning a study that will apply SJT to legal education in England. The target population will be training officers and senior magistrates, and the study will focus upon the identification of differences between participants' expressed emphases upon cues and their actual behavior in a judgment task dealing with such distal variables as trainer competence and severity of sentence.

Decision Support Systems Examined

Len Adelman
Department of Systems Engineering
George Mason University

We've continued research examining the effect of (a) causal focus on option generation, and (b) information order on expert judgment; in the latter case, on that of Patriot air defense officers. We've also completed a project investigating the types of decision support systems that are successfully adopted for operational use in the Navy. To put it bluntly, they are simple procedural or calculational aids, not sophisticated AI or decision-theoretic systems. And we have been actively involved in a cognitive systems engineering effort to design and evaluate alternative displays for AWACS.

Lens Model Used to Capture Group Judgment Policies

Reid Hastie, Center for Research on Judgment & Policy
University of Colorado, Boulder

Daniel Gigone and Reid Hastie used the lens model to capture the judgment policies of groups of college students who estimated the grades received by other Introduction to Psychology students. Each group (and its members individually) made 32 judgments and standard lens model statistics were used to summarize the impact of six cues (e.g., the to-be-judged student's attendance record, high school GPA, etc.) on the group and individual judgments. The focus of the study was on the effects of the pattern of information distribution across members of the group. A "Common Knowledge Effect" was observed such that the more members of the group who knew an item of information, the greater the influence of that information on the group judgment. Additional regression analyses were conducted to evaluate some alternate models for the process by which the "Common Knowledge Effect" was mediated. Gigone and Hastie would be grateful for any references to other studies that have used the lens model approach to capture the judgment policies of groups (rather than individuals).
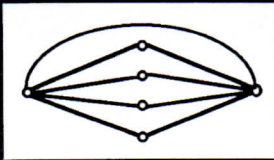
Research in Uppsala Continues to Focus on Problems of Dynamic Decision Making

Berndt Brehmer, Department of Psychology, Uppsala University, Sweden

The research in Uppsala continues to have its main focus on problems of dynamic decision making. We do both basic research and applied research in the area. However, we also continue to do more traditional Brunswikian research, using policy capturing to study the decisions by dentists.

Laboratory Research: Dynamic Decision Making, Distributed Decision Making

The laboratory research is conducted with *microworlds,* that is, computer simulations of complex systems. The most popular systems are NEWFIRE, a PC version of our fire fighting task, and MORO, a system developed by Dörner and his associates in Bamberg. We have also developed a system which enables us to study distributed decision making. This system is called D³FIRE and represents the fire fighting task on the level of the individual fire fighting unit, and makes it possible to study how groups of up to four subjects co-ordinate their actions to perform a common task (such as fighting a fire).

NEWFIRE is written in Smalltalk and requires a 386 PC with a 387 coprocessor and 2Mb of RAM and EGA graphics (and a Smalltalk license). D³FIRE is written in C and requires four PCs (286 or better) with 1 Mb of RAM and VGA color graphics which may communicate via a net or via their serial interfaces. MORO runs on any PC, but requires an experimenter who knows German.

Recent experimental work has been concerned with:

**1. How subjects learn the *time constants of dynamic systems.* **The

time constant is a special form of feedback delay caused by the fact that a system requires time to respond to a command. The time constant of a system is a well-known source of problems for process operators. Our experiments here have been conducted with NEWFIRE, and they show that subjects do not adapt to the time constants, and that they seem to develop strategies that help them avoid having to learn these constants.

**2. The effects of introducing *costs for actions in dynamic tasks.* **These experiments have also been conducted with NEWFIRE. The first experiments

show that introducing costs have negative effects on performance in a dynamic task. The explanation seems to be that when costs are associated with an action, subjects become less willing to experiment to find a good strategy, and the learning of the structure of the dynamic system is therefore slowed down. If, on the other hand, costs are introduced only after a period of practice without costs, there are indications that they may have a positive effect.

**3. Effects of *various forms of practice.* **From our earlier experiments, we

Risk Communication

Tim Earle and George Cvetkovich
Western Institute for Social & Organizational Research
Western Washington University

The focus of our work for the past several years has been risk communication. Progress in this field has suffered from the lack of a unifying theory. During the past year we have happily made what we think is useful progress in that direction. Our framework is centered on a conception of social trust as a risk judgment. The function of social trust (and its functional equivalents such as distrust, control, retreat, etc.) is to reduce cognitive complexity. In contrast to traditional views which base social trust on competence and responsibility, our interpretation uses a basis of cultural values that vary across persons, contexts, and time. In addition, we identify two types of social trust, normal and creative. The former is based on existing values, the latter on new values created through a process of social interaction. We use creative social trust as the foundation of a new form of public participation that we believe will prove useful in the solution of societal disputes in environmental management. For those with the interest and time to read a lengthy description of our theory, a draft version *(Social Trust: Pragmatic Risk Communication)* is available on request.

Uppsala Research

know that subjects do not develop very efficient strategies in dynamic tasks. We are examining the effects of new forms of practice that have been developed as part of NEWFIRE. These procedures make it possible for the subjects to see optimal strategies for a set of fires (it is not possible to develop a general optimal strategy, but optimal procedures can be found for certain fires). The results so far have been disappointing.

4. Effects of *specific and general goals* when controlling a dynamic task. These experiments are conducted with MORO, a system which requires the subject to assume the role of an adviser to a tribe in south Sahara with the general task of improving the conditions for this tribe. To perform well in this task, the subject will have to translate the general goal into a set of specific goals. MORO is therefore suitable to studying the goal formulation process, a very important aspect of performance in complex dynamic tasks. When we analyze subjects' performance in a system like MORO, it is clear that successful subjects are also those who formulate adequate goals. However, giving subjects adequate goals to start with does not help them, presumably because of the close relation between the goals and the mental models of the task; if the subjects lack an adequate mental model which makes sense of the specific goals, they are not helped by these specific goals, nor do specific goals seem to help the subjects develop adequate mental models.

5. Studies of *pathologies of decision making.* Earlier studies by Dörner and his associates have shown that subjects who perform badly in complex dynamic tasks exhibit a number of pathologies of decision making, that is, a number of characteristic errors. There has been some disagreement concerning whether these pathologies are precursors or consequences of failure, however. We have therefore chartered the occurrence of these pathologies over time, and shown that they are the precursors of failure, at least in a system such as MORO.

6. Studies of *heuristic and epistemic competence.* Heuristic competence refers to a general competence for handling complex dynamic tasks, while epistemic competence refers to domain knowledge about the system which one seeks to control. The pathologies mentioned under "5" above may be seen as a sign of deficient heuristic competence. In our experiments we have tried to induce both heuristic and epistemic competence by means of training and instructions. The results of these studies are currently being analyzed.

7. Studies of *strategies in distributed decision making.* Subjects may coordinate either in a *feedback mode,* using communication or in a feedforward mode, by developing and following a plan. We have found evidence of both these strategies in our data. We are now setting up experiments to investigate how the subjects' strategies for coordination vary with task characteristics.

8. Studies of *what is required for coordination.* The question here is: What do the subjects have to say to each other to achieve coordination? Our hypothesis is that they need to communicate their *intentions,* for it is only by knowing the intentions of another subject that it becomes possible to adjust one's own behavior to that of this other subject. Our evidence seems to support this hypothesis.

9. Studies of different *architectures of distributed decision making.* In D³FIRE, it is possible to create different architectures of distributed decision making by controlling who can communicate with whom. Our experiments so far have mainly been concerned with two kinds of architectures; a democratic architecture where each subject can communicate with every other subject directly, and a hierarchical architecture, where all communication has to pass through one centrally placed subject. The latter architecture is, in principle, better suited to the fire fighting task for it makes it more likely that all the relevant time scales of the task will be attended to.

10. Studies of *different forms of communication in distributed decision making.* In D³FIRE, communication is normally by means of "electronic mail." We have compared this to ordinary, face-to-face verbal communication, and found no important differences.

Applied Work: Decision Aids for Process Operators, Studies of Decisions by Dentists

Our applied work is concerned with decision aids and information systems for operators in pulp plants. The aim is to find ways of displaying the information that is stored in the information systems in these plants in a way so that the operators can learn from it.

A different line of applied research is concerned with decisions by dentists to extract or not extract teeth. This uses policy capturing methodology where we use X-rays as information together with other forms of information about the patient. The subjects, who are practicing dentists and oral surgeons, make a decision for each tooth—whether it should be extracted or not—and judgments about the strength of the indication that it should be extracted, as well as about a variety of problems that may occur and which form the basis for the decision to extract. Thus, we have a multistage problem, starting with the X-ray and the other information about the patient via the perceived cues in the form of possible problems and their likelihood and the overall rating of the strength of the indication that the tooth should be extracted to the actual decision. This raises a number of interesting modelling problems, with which we are now struggling.

Papers are available in English on most of the subjects described above. If you are interested, write to me or send me an e-mail message. The address is Uppsala University, Department of Psychology, P.O. Box 1854, 751 48 Uppsala, SWEDEN, e-mail PSYBB@seudac21.bitnet.