Table of Contents/Research Summaries

Integrating Judgment Research with Other Areas of Cognitive Psychology

Peter Juslin
Uppsala University

The research group in Uppsala on judgment under uncertainty is currently directing its attention towards integrating research on judgment with research in other areas of cognitive psychology. This involves both the following up of old lines of

research and the initiation of new ones. We are continuing our work with the sensory sampling model (Juslin & Olsson, 1997, *Psych. Review*, 104, 244-366), and we are currently about to finish or revise a couple of papers on these topics: One applies the model to auditory discrimination (loudness and pitch); one concerns the effects of feedback on sensory discrimination with pair-comparisons and single stimulus tasks (which differ); one concerns issues of mental speed and IQ; and one applies the model to dynamic event perception (written with Sverker Runeson). This autumn we will also run an experiment that provides a broad comparison of "Thurstonian" and "Brunswikian" uncertainty, in a number of respects not investigated previously.

The research on hindsight bias culminates with the doctoral dissertation of Anders Winman, which will be available in December 1997; some of its contents are also presented in Winman, Juslin and Björkman (in press; JEP:LMC).

We are currently finishing a paper concerned with a theory of how people initially partition their probability spaces when they make subjective probability assessments-the theory of Personal Probability Spaces (PPS). In a different line of research, we have tested a novel explanation of the base-rate inverse and base-rate neglect phenomena in inductive categorization tasks, an explanation that does not seem inconsistent with the exemplar-based models of categorization (these phenomena have seemed problematic for these models). We hope to have a paper on this "elimination hypothesis" ready by the end of the year. Finally, we have worked on a couple of chapters for a book that summarizes Swedish JDM research, where one of the important traditions represented in the book is the one inspired by Brunswikian psychology. One of the motivations for the book is, indeed, the fact that the researcher who initiated Swedish JDM research and the Brunswikian tradition in particular-Mats Björkman-is celebrating his 70th birthday this year.

Contact the author   Back to Table of Contents

Reason and Emotion

Kenneth Hammond
University of Colorado

Since finishing *Human Judgment and Social Policy*, I have been working on a second book manuscript that focuses on the struggle between reason and emotion as that struggle is involved in professional, moral and political judgments. Pursuit of this topic has led me to places I had never imagined I would go, topics I had never imagined that I would encounter, and best of all, surprises in what I have found. My main surprise has been to discover how much my education in judgment and decision making-albeit sadly lacking in the view of many-has enabled me to see aspects of these places and topics that others, more familiar with them than I, have not. I will describe a few of my surprises in the fields of history, government and law. (For those who think that they may be bored, I promise tales of illicit sex in the highest of places, government places, that is.) I am now confident that historians, political scientists and legal scholars have much to learn from students of judgment and decision making, and it will be my purpose to encourage members of the audience to teach them.

Contact the author   Back to Table of Contents

Social Trust

Timothy Earle
George Cvetkovich
Western Washington University

In the past year, George Cvetkovich and I have taken our social trust studies in a more explicitly Brunswikian direction by developing an adaptationist, biological account. This consists of an outline of the evolution of trust and a description of the connections between this biological heritage and modern social trust. Trust is the innate basis for all forms of human sociability. It begins with the parent-child bond and extends outward from here/now to include, first, kin and then non-kin members of small, local groups. Social trust evolved as a cultural response to the modern need to extend trust in time and

place and across group boundaries. This expansion of social trust can be facilitated by bracing moves toward inclusion in the solid foundation of innate trust and narrativity. This, in briefest summary, is naturalized social trust.

We identify two forms of social trust, pluralistic and cosmopolitan. Pluralistic social trust is singular, rooted in the pasts of existing groups. Since it is a within-group phenomenon, pluralistic social trust is not useful in the management of complex societal problems. Cosmopolitan social trust is multiple, created in the emergence of new combinations of persons and groups. These new combinations are based on new sets of values that are constructed for the solution of specific societal problems. The general public policy problem is this: How do you extend social trust across groups? We offer five preliminary steps:

- 1. The first step is to note that both forms of social trust are cultural adaptations to specific problems of life in the modern world. In contrast to the trust that evolved in the era of evolutionary adaptedness, social trust developed as a means of enabling people to take advantage of complex social relations by transferring control over specific circumstances to persons or institutions they could not know well.
- 2. The mechanics of pluralistic social trust are based on the operations of an innate "trust module." Thus, judgments of pluralistic social trust are made non-consciously; they are intuitive beliefs.
- 3. Judgments of cosmopolitan social trust are made consciously; they are controlled primarily by communication across existing group boundaries. Judgments of cosmopolitan social trust are reflective beliefs.
- 4. Our contemporary social trust problem is this: pluralistic social trust dominates cosmopolitan social trust. That is, the spread of cosmopolitan social trust is blocked by intransigent pluralistic social trust.
- 5. Proposed solutions to the social trust problem must be created from the materials at hand. Thus, rather than ignoring or directly opposing pluralistic social trust, proponents of cosmopolitan social trust should base their communications on innately-supported narrative structures.

Not all of our work is speculative. We have completed an interview study in which narrative structures related to health were compared among four groups of subjects: health care professionals, environmentalists, health-concerned individuals, and non-involved individuals. A report outlining the results of this study will be available in October. Our next study, to be conducted this fall, is an experimental study designed to explore the notion that judgments of pluralistic social trust are made non-consciously.

Our paper on naturalized social trust is available now. Request "Social Trust in Context: Biology, Culture and Public Policy" from:

- Tim Earle
- Psychology Department
- Western Washington University
- Bellingham, WA 98225

Additive and Explanation-Based Reasoning Displays

Leonard Adelman
George Mason University

When doing our research studying order effects with Patriot air defense officers, we found that an additive display presenting cognitive feedforward regarding the relative weights that the Patriot algorithm used did not induce the officers to act more additively and, consequently, did not eliminate officers' order effects when identifying aircraft as friend or foe. We hypothesized that this occurred because, in many cases, (a) officers were using explanation-based reasoning (EBR) when making their identification judgments and, therefore, (b) that an EBR display that explained-away cue values that were inconsistent with the additive model's recommendations was needed to induce officers to act more additively. In short, that the display needed to be consistent with how the officers were processing information.

Although we were unable to continue our research with Patriot officers, my students and I have now conducted two experiments with college students using a Patriot-like, hypercard-based simulation. We have found the following: (1) when participants were trained to use pattern-matching behavior (and EBR to understand the patterns), an EBR-based display that

explained away information that contradicted the recommendation of an additive model induced more additive judgments than an additive display; and (2) when participants were trained additively, EBR and additive displays were equally effective, thereby suggesting the overall superiority of an EBR display. However, (3) the effectiveness of the EBR (or additive) display was highly dependent on the characteristics of the particular situation. This finding is of particular concern for system design because difficult judgment situations are typically unanticipated by the operator and designer alike. Our most recent experiment investigated whether the availabity of an additive display improved the EBR display's robustness for difficult judgment situations. Initial data analysis suggests that it failed to do so.

Generalizability Theory and the Lens Model Equation

Steve Schilling and Jim Hogge
Vanderbilt University

We are at work on a paper exploring how hierarchical linear models (Byrk & Raudenbush, 1992) can link two distinct analytic models of judgment: generalizability (G) theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) and the lens model equation (Cooksey, 1996) of social judgment theory (Brehmer & Joyce, 1988).

The need to deal with the dependability (reliability) of judgments arises naturally in the analysis of judgment. For example, if raters have been asked to assess the competence of individuals on the basis of several cues, it would be of interest to estimate the dependability of those judgments (1) over time and (2) across raters. From the point of view of G theory, each judgment is considered to have been sampled from a multifacet universe, in the present example comprised of facets for occasions, raters, and persons. The primary analytical tool of G theory is the analysis of variance, which is used to estimate multiple sources of error (variance components). While G theory identifies specific facets of the judgment environment in order to permit specification of the conditions that the data user would be equally willing to accept (e.g., generalizing across raters and occasions), these facets are not explicitly used to model the judgments. Viewed in this light, G theory seeks to overcome diversity in conditions and raters for the purpose of identifying a single universe score that can then be applied across all relevant conditions.

In contrast to G theory, the lens model equation explicitly uses cues in the judgment environment to form a linear regression model of judgment for each rater. A major accomplishment of the lens model equation is its ability to decompose lack of agreement between two raters and to model incompatibility between raters and inconsistency of model application within raters. The primary analytical tool associated with the lens model equation is multiple regression. The chief objective of the lens model equation is to form workable models of individual judgment processes. The lens model equation implicitly accepts diversity in judgment by using a separate model with differential weighting of cues for each rater.

Given the differences in orientation, objectives, and methods between these two approaches to analyzing judgments, one framework seems to have little implication for the other. We will show that the opposite is true. G theory has profound implications for the lens model equation in that values of variance components in G theory act to constrain possible values obtained in the lens model equation. However, the central obstacle to a useful fusion of the two approaches is the failure of G theory to incorporate cues into its variance decompositions. Extending G theory to incorporate cues naturally leads to the analytic methods of hierarchical linear models, where random effects of cues across raters are explicitly modeled in terms of variance components.

Specifically, we will show how the linear variance component models used in G theory can be seen as a subset of hierarchical linear models (HLMs). Therefore, the first extension of G theory within the context of HLMs is the incorporation of a common linear regression model across raters, which is demonstrated to be explicitly aliased with the persons variance component in G theory. The second extension-decomposition of the persons by raters variance component into random effects of cues and error-will be shown to yield the random regression HLM. Further extensions will be shown to yield HLMs incorporating occasion effects. Finally, we will demonstrate how diverse characteristics of individual raters can be used to model the diversity of judgment across raters.

These connections will be illustrated with data from two empirical studies. The first was an investigation of ratings of the professional competence of hypothetical nursing students by nursing faculty and staff (Hogge & Murrell, 1994). Nine members of the faculty of a hospital-affiliated nursing school in London, England and 20 hospital nurses responsible for the formative assessments of the practice of advanced students from the same nursing school rated the overall competence of 25 randomly-generated hypothetical nursing students whose performance with respect to seven performance criteria had been

summarized on profile sheets. The same raters later rated the overall competence of the same 25 hypothetical students. The second study dealt with expert forecasting of hail (Stewart, Moninger, Grassia, Brady, & Merrem, 1989). Seven meteorologists used six items of information derived from Doppler radar volume scans of 25 storms to make probability forecasts of hail and severe hail on two occasions.

These empirical applications will show how the use of HLMs links G theory and the lens model equation and extends both as follows:

- 1. HLMs provide a means for estimation and testing of the common aspects of a model across raters as well as modeling variability between raters, all within a coherent methodological and statistical framework.
- 2. HLMs extend the lens model equation by permitting decomposition of lack of agreement between raters into error variance and random effects of cues.
- 3. Sources of unreliability that can be examined are extended beyond a standard G theory analysis (e.g., the impact of rater's differential cue utilization) and can be taken into account in the design of subsequent decision studies.
- 4. Estimation of the individual regression models by empirical Bayes methods within the context of HLMs is shown to be more reliable than the least squares estimates commonly used in connection with the lens model equation.

In summary, HLMs are a promising analytical tool for combining two previously divergent approaches into a single coherent framework-a framework simultaneously capable of assessing the dependability (reliability) of judgments and modeling diversity in judgment models across raters. In effect, G theory is extended to take into account the information (cues) upon which judgments are based, and the lens model equation's pairwise consideration of interrater agreement is supplanted by an analytic framework that considers all raters simultaneously, both individually and collectively.

References

Brehmer, B., & Joyce, C. R. B. (Eds.). (1988). *Human judgment: The SJT view*. Amsterdam: Elsevier.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park CA: Sage Publications.

Cooksey, R. (1996). *Judgment analysis: Theory, methods, and applications*. San Diego: Academic Press.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles*. New York: John Wiley.

Hogge, J. H., & Murrell, J. (1994, November). A look at nursing competence through Brunswik's lens. Poster presentation at Annual Meeting of Society for Judgment and Decision Making, St. Louis MO.

Stewart, T. R., Moninger, W. R., Grassia, J., Brady, R. H., & Merrem, F. H. (1989). Analysis of expert judgment in a hail forecasting experiment. *Weather and Forecasting*, **4**(1), 24-34.

Brunswik-Symmetry

Werner W. Wittmann, Universtät Mannheim

This has been a busy year where we tried to survive cuts in the funding of our University and, fortunately, we were successful. The year was also, in many aspects, devoted to popularizing Brunswik-Symmetry in Germany and elsewhere. I've been invited by Phil Ackerman, Pat Kyllonen and Rich Roberts to an international conference: "The future of learning and individual differences research: Processes, traits, and content," to be held at the University of Minnesota on Oct. 9-12, 1997. The title of my presentation is: "Investigating the paths between working memory, intelligence, knowledge and complex problem-solving performances via Brunswik-symmetry." Those interested in that conference can find information under: http://www.brooks.af.mil/AL/HR/HRC/HRCC/minnconf.htm .

I was also invited to the ISSID97 (International Society for the Study of Individual Differences) at Arhus, Denmark, to contribute to a symposium about new directions in ability research. The title of my presentation was: "Challenging g-mania in intelligence research: Answers not given, due to questions not asked." As you can imagine, the questions not asked are related to not using the lens-model and the Brunswik-symmetry framework. I demonstrate in that paper how the US Army and many others are fooling themselves if my assumptions are true and if my results from Germany are generalizable to the States. You can look at that handout/paper at our homepage under: http://zombie.psychologie.uni-mannheim.de/psycho2/forsch/papers/papers.htm, and you'll find the reasoning still in my bad English, but also a bunch of very nice figures and numbers. Those of you who have seen my figures at Chicago will probably have a deja vu experience.

At the same URL, you'll find a paper which Tom Cook (90%) from Northwestern University and I (10%) have written for a European audience concerning evaluation research. In the subheading "more and better evaluation theory," we intensively pointed to the importance of Kenneth Hammond's new book. I personally strongly believe that evaluation researchers can only profit from Ken's book and his and others' work on social judgment theory (as I force my students to do).

At our homepage. http://zombie.psychologie.uni-mannheim.de/psycho2/forsch/reports/reports.htm, you will find our research reports series, some of them in English. For Brunswikians (Heft 8), which means number 8, should be of interest-see the title mentioning Brunswik-symmetry. All reports are downloadable in adobe acrobat or postscript format. The homepage also contains some descriptions of our research program in English. We promise to make the navigations for English-speaking visitors more comfortable in the near future.

Decision Conferencing-A Fair Process?

Franklin Laufer
University at Albany

In the fall 1996 issue of the Newsletter, under the heading "Effectiveness of Decision Conferencing," I described the work John Rohrbaugh and I initiated to answer the question, "Can decision conferencing ensure fairness in group decision making?" Our survey of participants in three conferences, including facilitators, panelists, and observers, and analysis of responses was completed. Despite the limited number of conferences included as the objects of evaluation, the following results or findings can be reported: 1) a set of seven summative scales was developed to measure participants' perceptions of adherence to the procedural justice criteria or rules proposed by Leventhal (1980), corresponding to aspects of the decision conference process (representativeness, ethicality, correctability, accuracy, and bias suppression) as well as to the resultant policy model (accuracy and consistency); 2) respondents generally agreed that the decision conference process was procedurally fair and adhered to procedural justice rules, especially representativeness, ethicality, correctability, and bias suppression; 3) differences between conferences in panelists' perceptions of adherence to the rules of representativeness and accuracy of the model were found, suggesting the importance to perceived fairness of the process of who participates as well as the accuracy of the criteria contained in the final policy model and the ability to measure these criteria; 4) across all conferences, the correlation between assessments of overall process fairness and correctability was the highest of any of the correlations between each of the procedural rule scales and overall fairness, suggesting a strong and consistent relationship between panelists' perceptions of procedural fairness and correctability; 5) facilitators generally provided higher ratings of adherence to procedural rules than panelists, with observers' ratings typically intermediate.

We evaluated the implications of these results for participation in resource allocation decision making. We concluded that procedural justice can provide a comprehensive framework for assessing the fairness of participation, especially when implemented through decision conferencing, in the formulation of policy.

Aging, Probability Learning, and the Detection of Invalid Cues

Gérard Chasseigne
Université Francois-Rabelais

The effect of aging on the ability to discount invalid cues in the process of learning probabilistic relations has been studied. The discounting of information constitutes an important cognitive function. In order to function adequately in our environment, we must be able to select from a myriad of stimulations the (generally) few relevant ones and ignore those which are not critical to our current goals.

Generally, people have difficulties in identifying and separating irrelevant information from relevant information in making judgments. Hasher and Zacks have proposed that age differences in cognitive performance are a consequence of greater vulnerability to interference or distraction. Older people should essentially be less efficient in inhibiting irrelevant information.

Several authors, indeed, have found that older adults have more difficulties than young adults at discounting irrelevant material in a variety of tasks: reading comprehension, medical information processing, maze learning, Stroop-like tasks. In all these studies, discounting invalid cues appeared as an active process. In other words, inhibition consumes cognitive resources.

A total of 48 individuals (16 in each of three age groups: 20-30 year-olds, 65-75 year-olds and 76-90 year-olds) participated in a multiple-cue probabilistic learning experiment. None was institutionalized. The three groups were comparable in terms of years of formal education. All participants had a sufficient vision to perform the task.

The material consisted of three sets of 26 cards each showing four cue values in the form of vertical colored bars noted A, B, C, and D, whose heights varied from card to card. The criterion value was written on the back of each card. The ranges of the three bars were the same. The cue distribution was normal. Sets of cards were constructed so that the cue intercorrelations were zero. Cues A and C were distracters; the correlation between cue A or C and the criterion was .00. The correlation between cue B and D and the criterion was always .68. The square of the multiple correlation between the criterion and the three cues (R^2) was .92.

The participants were told that the task which faced them was a weather forecasting task. They were asked to learn the relationships between the levels of the four indicators and the pleasantness of the next day's weather by means of different situations which would be provided, each of which was characterized by sets of four cue values displayed on the front of different cards and the actual value of the criterion displayed on the back of the same cards (outcome feedback, OFB). The subjects were also told that an exact weather forecast was nearly impossible because of a myriad of other factors acting independently. They were finally told that some cues were more important than others. The experiment was self-paced. The subjects were shown six blocks of 26 trials. The first three blocks were those described previously and the last three consisted of the former with an inverse order presentation. No OFB was provided during block 1. This block served as a reference. OFB was provided from block 2 to the terminal block.

Utilization of cues A and C constitutes the central results considering our hypothesis, that of greater difficulty in elderly to detect invalid cues (Age x Block interaction). For cue C, the 20- to 30-year-olds' utilization values started high, with a mean value close to r =3D .45. They then decreased to .20 (second and third blocks) and finally tended to zero. The 75- to 90-year-olds' utilization values started much higher (.70). They then decreased to .20 (second and third blocks) and remained at this level. A similar pattern of results was observed as concerns cue A utilization. In elderly people, cue utilization in the terminal block was always significantly different from .00.

Our hypothesis was thus well supported by data. Only the 20- to 30-year-olds as a group succeeded completely in not using the invalid cues at the end of the experiment. In the two other groups, cue A or C utilization was decreasing at the end of the experiment, but nevertheless different from zero at the terminal block. (Verbal protocols about cue utilization were in agreement with quantitative results.) Older adults were thus less flexible than younger adults in abandoning ineffective strategies.

Inner Ears, Emergency Rooms, and Public Policy

Jeryl Mumpower, University at Albany

Several papers describing research in which I was involved with other colleagues were finished this year. A paper on physician's diagnostic judgments and treatment decisions for infants with inner ear infections, on which Claudia Gonzalez-Vallejo took the lead and several others of us collaborated, will be appearing in Medical Decision Making. We also finished the first of what I hope is a series of papers describing our research on judgment and decision making in psychiatric emergency rooms. My student Bruce Way, who just earned his degree for all his good work, is the lead author on this paper, which reports on the level of inter-rater reliability among eight psychiatrists who viewed and rated 30 videotaped intake interviews. For most of the pertinent judgments, we found relatively low levels of agreement, either with the assessing psychiatrist in the emergency room or with one another. The eight psychiatrists agreed best on judgments of psychosis (.68) and levels of substance abuse (.67), but they exhibited low levels of agreement regarding, for instance, degree of psychopathology (.28), impulse control (.30), and danger to self (.35). They even disagreed among themselves about the appropriate disposition (.34).

The project that I am most excited about right now is a collaboration with Tom Stewart. We are planning to try to apply the key ideas in Ken Hammond's Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice, to contemporary policy problems (including ones we are already working on such as breast cancer screening, decisions in psychiatric emergency rooms, and global climatic change). We hope that this will help to make sure that Ken's important ideas are more widely understood and appreciated in the policy world.

Brunswik Down Under

Ray Cooksey
University of New England

My Brunswikian research is evolving on two fronts, one more traditional and the other much more systemic and radical in nature (therefore much more embryonic at this stage). On the traditional side, one of my Ph.D. students and I are commencing an investigation into the effects of initial stage and middle stage Alzheimer's disease on decision making. We are adopting a cognitive continuum approach to exploring the capacities of Alzheimer's patients to make various types of analytical and intuitive decisions. Preliminary work has been done in this area by Damasio and associates, among others, but a more coherent (and correspondent) approach is needed. Some of this work invokes Epstein's concept of experiential and rational cognitive systems which shares common features with cognitive continuum theory. We will be exploring patient capacity to solve everyday problems (such as personal financial management, navigation from point to point, choice making such as activities to engage in, groceries to buy, or clothes to wear, judgments of emotional states of others) which vary along the task continuum. Some predicted outcomes include: more advanced Alzheimer's patients should show stronger decrements in capacity to make decisions of an intuitive or emotional nature because of the areas of the brain which are affected (decreased correspondence or accuracy) and should show greater problems in achieving congruence between mode of thinking and task demands (e.g., persistence in relying on intuitive judgment, however poor, even after learning of the existence of analytical strategies which produce correct solutions to problems).

My other area of research endeavor is not strictly Brunswikian, but incorporates many of Brunswik's methodological concepts into a larger systemic framework which focuses on understanding human behavior (including decision making and job performance) in organizational contexts. This is a generalized extension (beyond the Lens Model) of the perspective I presented at the Brunswik Society Meeting two years ago. We are trying to synthesize a large and diverse body of literature and contrary research findings (most of which were nomothetically established) into a holistic and systemic approach to understanding work performance. The perspective is distinctly idiographic, nonlinear, and dynamic, involving multiple positive (destabilizing) and negative (stabilizing) feedback loops between various subsystem components (environment and task, organization, group, and individual). What is Brunswikian in this framework is the methodological focus on multiple tasks, multiple outcomes and multiple contexts for a single individual and analyzing behavior through time. The cognitive continuum (in conjunction with aspects of Epstein's perspective) also plays a prominent role in this approach. The framework also integrates quantitative and qualitative methodologies to provide a triangulated approach to the study of behavior. We want to test some initial propositions from this perspective in the context of legal decision making by Australian magistrates and court justices (many of whom operate in the context of large public or private sector practices). [A preliminary test of the fundamental interplay between the various subsystem components invoked in the theory (in an educational context) has shown a strong but variable component of chaotic dynamism through time for individual workers.]

Why Do Doctors Disagree When the Evidence Is Clear?
A Study of the Impact of Clinical Trials
and Conflicts of Values

D. Mark Chaput de Saintonge, Maria Woloshynowych
St. Bartholomew's and the London Hospital Medical Colleges
Roy M. Poses, Memorial Hospital of Rhode Island

Angiotensin converting enzyme (ACE) inhibitors are drugs proven by randomized controlled trials to extend life and reduce symptoms for patients with congestive heart failure (CHF). There is good evidence that physicians fail to prescribe these drugs for many patients who would benefit from them. The reasons for this apparently paradoxical phenomenon are unknown. Our hypothesis is that physicians may avoid prescribing drugs for individual patients whom they judge to be less likely to benefit from the drugs or more likely to have adverse effects from them. Furthermore, their judgments of the probability of benefits and adverse reactions may be based on cues that in fact do not predict these outcomes.

We are using clinical judgment analysis to identify cues used by physicians to make these judgments. The study is recruiting a random sample of General Practitioners, Internists and Cardiologists from the London Metropolitan health area. Vignettes describe patients for whom an ACE inhibitor would be indicated. For each vignette, we ask physicians to judge: the likelihood they would prescribe an ACE inhibitor; survival after one year were an ACE inhibitor to be prescribed and survival after one year were no ACE inhibitor to be prescribed (allowing computation of their judgments of the survival advantage due to ACE inhibitors); and the probability of adverse drug reactions were an ACE inhibitor to be prescribed. We are also surveying these physicians about their general propensity to prescribe ACE inhibitors and their judgments of the rates of the outcomes of interest in the general population of patients with CHF. Cues include factors known to predict at least one of the outcomes of interest. We also chose as distracter cues factors that in some sense resemble predictive cues but are not themselves predictive, and factors that relate to disease severity in general. We are using a fractional factorial design modified to exclude biologically implausible cases.

Some subjects have a lot of difficulty with the task, especially with what they claim to be their lack of knowledge of absolute figures for survival rates. Preliminary analysis (of 12 cases) so far shows relationships between some physicians' likelihoods of prescribing an ACE inhibitor and their judgments of survival advantage due to ACE inhibitors and of rates of severe adverse drug reactions.

Patient Judgments About Anti-depressants

Celia E. Wills
Michigan State University

This year I have collected data for two studies.

One was a study of how college student volunteers (N=89) made judgments of likelihood of taking a hypothetical antidepressant medication, given (nonvarying) efficacy information and varying information about risk of nausea, presented in positive/negative frames and varying the format of base rate information (absolute % risk; absolute number risk; base rate/100,000; and base rate/100). Students who made ratings for only one type of frame were more likely to have zero variability in likelihood and confidence ratings than students who made ratings for both positive and negative frames. Likelihood and confidence ratings were positively correlated. An abstract for this study has been submitted for review for a poster presentation at JDM.

The other study used a community-based convenience sample of adults (N=96) who recently (less than four months at time of survey completion) had made a decision about whether or not to take an antidepressant medication that had been suggested or prescribed by their primary care health care provider. This study did not include a judgment task, but focused on assessing the feasibility of using standardized measures (desire for decision participation, decisional conflict, satisfaction with health care provider) with this population. Approximately 23% of the sample reported declining or prematurely discontinuing medication (within less than four months of starting). Participants who declined medication reported a significantly greater preference for control in decision making, had greater decisional conflict, and less satisfaction with their health care provider compared to those who had started medication. Participants who discontinued their medication reported significantly greater decisional conflict compared to those who were still taking medication. An abstract for this study has been accepted for a paper presentation at the Annual Meeting of the Midwest Nursing Research Society in April 1998.

Prediction of Patient Follow-up

Robert M. Hamm
University of Oklahoma Health Sciences Center

With colleagues Kathryn Reilly, MD, and Vickie Loemker, MD, during the coming year I will do a study (funded by the Cancer Research Foundation of America) of patients attending a women's clinic for cervical dysplasia. The data can be viewed in a Lens Model framework.

The study is pertinent to the decision about how mild cervical dysplasia should be managed once it is diagnosed with Pap smear and colposcopic exam of the cervix. The patient can either receive an immediate procedure (cryotherapy) or "watch and wait" in the hope that the dysplasia will cure itself spontaneously. The success of watching depends on the patient returning regularly for follow up exams. In the majority of cases, the dysplasia will regress. But if the dysplasia is left in place, and not checked regularly, it can progress to cervical cancer and be discovered "late," requiring a more extensive procedure, or even "too late."

The study will involve a sample of women patients and their doctors and nurses. The patients will fill out a questionnaire (working with a research assistant) that covers a number of factors we anticipate will be related to whether the patient will return regularly for follow-up. The nurse and physician will each make a prediction about whether the patient will return. The patient's success adhering to the follow-up schedule will then be measured.

The data fit the Lens Model framework as follows. The patient follow up behavior is the environmental criterion. The patient's answers on the questionnaire, plus whatever the nurse and/or physician observe, constitute the cues. The physician and nurse predictions are the judgments. Predicting the behavior from the set of cues is the environmental model. Predicting the judgment from the set of cues is the judgment model. Predicting the follow-up behavior from the nurse's or physician's judgments is the achievement.

Procrustes might complain about the fit of the data to the Lens Model framework. First, different nurses and physicians will be involved; some may see multiple patients, but it is not well controlled. A model of judgments of "physicians at this clinic" is not as clean as a model of an individual physician's judgment. However, once the data are collected, then profiles of real patients could be produced for later judgment by individual nurses and physicians. Second, the nurses and physicians will not have the full data set. They won't see the patient's answers on the questionnaire. It could be possible to have the nurse and/or physician tell us their perception of the cues-say, fill out the questionnaire as they think the patient would fill it out. However, this is not part of the current study design, and getting their cooperation in such a project may not be easy.

The purpose of the study is to determine if the patient's follow-up behavior is predictable a) by the physicians and nurses, or b) from any information that might be gleaned with the questionnaire. If patient follow-up is predictable, then it could be a basis for deciding how to treat the patient (if they won't follow up faithfully, then cure them with the procedure). If physician judgment is not accurate, but the objective data can be used to make an accurate prediction, then use it. If not predictable, then use the rate of loss of follow-up in the decision analysis and make a universal recommendation.

Testing Cognitive Continuum Theory

James Holzworth
University of Connecticut

My colleagues and I are continuing to do research in the Brunswikian tradition. Several new projects are underway. First of all, Steven Mellor, Jim Conway (Seton Hall University) and I are doing a judgment study investigating people's inclinations to be represented by labor unions. Employed persons (who are not currently members of labor unions) are being asked to make judgments concerning how likely they would be to vote in favor of union representation. We are hoping to find out what cue(s) people focus on when considering union representation. Eight cues are varied in each of 48 union representation scenarios. Four cues concern extent of potential disadvantages of union representation: unions can be antagonistic, unions

can be costly, unions can be exclusive, and unions can be corrupt. Four cues concern extent of potential advantages (benefits) of unionization: unions can provide a voice, unions can provide a grievance procedure, unions can provide a sense of job security, unions can promote respect and dignity. We are just beginning data collection.

Secondly, I am about to begin a series of studies to test premises of Cognitive Continuum Theory (CCT). Pilot testing will begin shortly, in collaboration with Janet Barnes-Farrell (industrial/organizational psychology), Judy Brown (medical cytogenetics), and Julia Pavone (fine arts). Study 1 is concerned with measurement of individual differences which might predispose people toward different modes of cognition. A battery of measures will be developed for assessing individual predispositions, including: biographical items concerning formal and informal exposure to science, math, and fine arts; tolerance for uncertainty; tolerance for ambiguity; and decision style. The test battery will be given to each participant in Studies 2-4, which are experiments designed to determine: (1) if different cognitive tasks induce study participants to employ different modes of cognition, (2) if participants oscillate along the continuum between analysis and intuition, and (3) if participants sometimes alternate between pattern recognition and use of functional relations. Studies 2-4 utilize within-subject designs, in which all participants serve in every experimental condition. Study 2 participants will be shown work samples of restaurant waitresses presented in several ways (videotapes, written transcripts, and summary data). Each participant will evaluate overall performance of the waitress and give oral justification. Study 3 participants will engage in two classification tasks: sorting photos of human chromosomes (all 23 pairs) into two sets (male/female) based on presence of a Y chromosome, and sorting photos into two sets (normal female/Turner Syndrome) based on presence of an anomalous X chromosome. Participants will "think aloud" while sorting. Study 4 participants will view different styles of art (representational and nonrepresentational), "thinking aloud" while viewing each painting. Participants will submit written comments and critiques of the art. In each experiment, verbal protocol analysis of "think aloud," justification, and evaluation data will be done to test premises of CCT. Multivariate analysis will be done on test battery data and performance measures to determine if participants' background variables correlate with initial use of a particular mode of cognition.

A third project concerns my interest in smart ridge regression (combining human judgment with ridge regression). Tom Stewart and I, hopefully with help from others, are slowly making plans to follow-up on my recent effort (OBHDP, December 1996). We hope to try out smart ridge regression in more and varied judgment tasks, beginning very soon. I hope to have some preliminary findings to discuss at our Brunswik Meeting in November.

Lens Model Studies of Calibration

Michael Doherty, Bowling Green State University

The fundamental focus in studies of calibration is how well subjective probability judgments match actual event outcomes. In this respect, research in calibration is very similar to Lens Model research in the Brunswikian tradition. In both lines of research, the focus is on empirical accuracy, or correspondence between judgments and the environment. Interestingly, calibration research has long been classified with the Heuristics and Biases program. While the current theories of overconfidence are greatly influenced by Brunswik (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Juslin, 1993, 1994; Björkman, 1994; Soll, 1996), the methodology employed in calibration studies is decidedly not, except with respect to random sampling from the environment.

Our calibration tasks are designed with Brunswik's strictures concerning representative design in mind, and conceptualized within a Lens Model framework as well as a calibration framework. This enables us to investigate empirical relationships among lens model statistics and calibration indices. In a JDM poster, we report the lens model reanalysis of calibration data on predictions of baseball games reported last year. In addition, we will have additional data on baseball judgments, as well as new data on a non-sports domain, the prediction of a roommates' judgments of desirability of other roommates. The new data are currently being collected by Greg Brake as his dissertation data.

Report from the Cognitive Engineering Laboratory

Kim Vicente, University of Toronto

During the past year or so, the Cognitive Engineering Laboratory at the University of Toronto has been involved in the following research projects:

1. After several revisions, we have finally completed a paper describing a novel ecological theory of expertise effects in memory recall. This paper will be published in the January 1998 issue of Psychological Review.

2. We have published a brief commentary in Human Factors (39, 323-328) whose goal was to make human factors researchers better aware of Brunswik's principle of representative design of experiments. A framework for human factors research based on this principle was described.

3. We have continued our work on ecological interface design (EID) by extending it into new domains, and by addressing several outstanding issues. For example, one paper (to be published in the International Journal of Aviation Psychology) applies EID to aviation, and another paper (to be published in the Journal of Clinical Monitoring) applies EID to medical equipment interfaces. Recent issues that we have been addressing include: the relationship between EID and individual differences (especially cognitive style), making the most of EID by instructing people to self-explain their actions, how the problem of interface navigation can be dealt with in the context of EID, and the development of new measures of performance and adaptation in dynamic microworlds. Finally, a review of our previous work on EID was published in the Systems Dynamics Review (12, 251-279), showing how interface design can mediate the success of dynamic decision making.

4. I am on sabbatical this year, and my main project is working on a textbook with Annelise Mark Pejtersen. The book will describe Jens Rasmussen's framework for cognitive work analysis, a set of concepts that can be used to analyze human work in order to derive implications for the design of computer-based support systems. The book is due at the publisher (Erlbaum) on May 1, 1998, so I will be hiding until then!

Contact the author   Back to Table of Contents

Organizational Fixes for Cognitive Limitations

Joshua Klayman
University of Chicago

While continuing to work on two "old" topics (judgments of confidence, with Jack Soll and Claudia Gonzalez-Vallejo and learning of causal systems, with Alex Wearing), I am also working on a new project with Chip Heath and Rick Larrick. We are looking at the possibility that organizations develop "fixes" for the cognitive failings of individuals. As a result, individuals in the organizational context may do better at a variety of cognitive tasks than they would if they were on their own. We see three main varieties of fixes: (1) social (using other people to aid judgment: e.g., the organization requires a proposal be presented to a meeting of representatives of different departments before being put forward); (2) provision of tools (e.g., training and requiring people to draw scatterplots, or requiring the "five whys" procedure to force people to look beyond surface explanations-"Why did the machine break down? Because the part broke. Why did the part break? Because of excess wear. Why was there excess wear? Because it wasn't lubricated properly. Why wasn't it lubricated properly?...."); and (3) conceptual (often captured in mottos, e.g., "Don't confuse brains and a bull market" to symbolize a pitfall of biased self-attribution).

So far, we are only trying to systematize some anecdotal data, but we think there is interesting research potential here. Given that most "fixes" seem to have been developed intuitively, or by trial and error, we wonder which shortcomings are more or less well fixed, how effective the fixes prove to be, and when attempted fixes may be ineffective or even counterproductive. Meanwhile, we welcome further anecdotal data from anyone who might have an interesting candidate fix.

Contact the author   Back to Table of Contents

Czech Republic

Lubomir Kostron
Masaryk University

- Helena Klimusova successfully completed and defended her M.A. thesis entitled "Analysis of Real and Artificial Data, Generated by POLICY PC." An English summary should be available soon.
- Martin Vaculik teaches how to use POLICY PC in group settings (John Rohrbaugh's automated conferences) of voluntary groups of psychologists both semesters and will continue to do so this coming school year.
- Martin Vaculik also submitted a paper about his experiences with teaching and implementing POLICY PC to the *Czech Psychologist*, our professional magazine.
- During October my own volume about you all will be published. The title *Psychology of Judgment-making: Theory and Methodology of E. Brunswik, Kenneth R.Hammond and Their Followers*. This book is the first psychological monograph in Czech about judgment and decision making.

Judging How One is Judged
by Others

Linda Albright
Westfield State College

My research continues to focus on interpersonal perception. One recent study concerns meta-perception (R. D. Laing, 1966), which refers to the judgment of how one is judged by others. In this study, we predicted that the ability to judge how one is perceived by others could be increased by putting people in the visual perspective of the observer. Small groups of four or five persons participated in a social interaction task while being videotaped. After the interaction participants judged each other on traits indicating the interpersonal competence dimension. Based on random assignment, groups then either did or did not watch the videotape of their group. The opportunity to observe oneself through videotape was the manipulation of visual perspective. After watching the tape (or a tape of another group-control condition), participants made predictions of how they were judged by the other members of their group (meta-perceptions). We found greater accuracy of meta-perceptions (greater correspondence between meta-perceptions and trait judgments made by others) among those who watched a videotape of themselves. Thus, visual perspective appears to account for some of the discrepancy between the way people think they are judged and the way they are actually judged.

Another study of meta-perception recently completed focuses on the mediational basis of meta-perception, or how we determine what others think of us. A recent paper by Kenny & DePaulo (Psych. Bull. 1993) outlined four models of the process of meta-perception. Briefly, the alternatives are: a) feedback from others; b) global self-perception; c) situational behavior as perceived by self; and d) situational behavior as self judges others would judge it. This study attempted to show that people can go beyond their judgments of themselves, and judge how others would interpret their behavior. To demonstrate this ability, participants were randomly assigned to enact a role while interacting with a person, who was not acting. By randomly assigning people to a role, the self-rating becomes uncorrelated with one's interpersonal behavior. Behavioral measures and interpersonal trait judgments were obtained. Results showed that actors' meta-perceptions were uncorrelated with self-ratings, but highly correlated with behavior.

Naturalistic and Experimental
Distributed Dynamic Decision Making

Alexander J Wearing, University of Melbourne, Mary M. Omodei, La Trobe University, Jim McLennan, Swinburne University

Our research focus over the last 18 months has been on multiple participant decision making under pressure of time. Our program has had two prongs. One prong has involved the participant observation of firefighters (in each situation, the commander) who themselves have worn a mini-TV camera with audio in their helmets to provide an "own point of view" record of their experience. Following the fire event, the participant observer (usually Jim McLennan) has debriefed the commander or incident controller. This debriefing has involved replaying the video and audio record in order to cue memory. (The "own-point-of-view" record has proved a very powerful cue). These qualitative data are facilitating the modification of those theories of decision making that have been developed to account for behavior in situations of this kind. Provisionally, factors that seem to be important are the decision criteria (what tasks, e.g., saving firefighters' lives, have priority on the fireground), the construction of a "working scenario" that is used to guide action, the employment of analogical experience in selecting the next steps, and the capacity to use feedback in adjusting the working scenario.

The second prong has involved the extension of FIRECHIEF (a computer based task that can represent oil spills, locust infestations, etc., but in our experiments is realized as a firefighting scenario) to allow distributed decision making. This has involved the networking of FIRECHIEF to allow multiple participants, each with their own sector of the fireground, and the development of a variety of command and control structures for linking the participants. To this point, we have examined three factors. First, we have investigated the effects of varying the amount of information available to, and varying the amount of control that is possessed by an incident controller (the commander) on the performance of the team. Second, we have looked at the effect on performance of allowing the orders of the commanders to be questioned as opposed to requiring that commands be obeyed without dissent. Preliminary findings suggest that, in these experimental conditions, providing more information and more control tends to degrade the performance of incident controllers (even though they see their workload as lighter, and possessing more information and control as being helpful), and allowing subordinates to provide dissenting feedback makes no difference to performance. Personal characteristics, cognitive variables, mood, and perceptions by the co-workers of one another's competence also seem to relate to performance.

Risk Ranking,
Organ Donation Preferences

Michael L. DeKay
Carnegie Mellon University

Since coming to Carnegie Mellon one year ago, I've been involved in a couple projects that might be of interest to Brunswikians (and several others that would probably not be of interest).

The first is a big project on Risk Ranking with Granger Morgan, Baruch Fischhoff, Keith Florig, Paul Fischbeck, and graduate students Karen Jenni and Kara Morgan. The ultimate goal of the project is to develop a defensible method for ranking risks to people and the environment so that government agencies like the EPA can set priorities in a more defensible fashion. Currently, the project team is focusing on risks to health and safety (but not the environment) in a particular test bed (a hypothetical middle school) and exploring how groups of laypersons and risk professionals rank these risks.

Each risk is described in its own booklet, which contains both qualitative and quantitative information, along with estimates of uncertainty where appropriate. The risks span the range of those encountered (e.g., asbestos, food poisoning, bus accidents, electromagnetic fields, sports injuries, etc.), so there is some natural correlation of risk dimensions across risks. Our procedure is exploratory and subject to periodic revision. A typical participant usually rates a subset of the risks (say 12) individually and in a holistic manner, before coming to the group. Groups of five to ten participants then attempt to reach consensus on the ranking of these risks. In some instances, participants (or groups) perform a multi-attribute utility task, so we will be able to compare the rank orders that result from intuitive (holistic) and analytical (MAUT) approaches. We may also use policy-capturing techniques on the holistic data and compare weight profiles and other measures of agreement across methods, subject groups, etc.

Preliminary results from this research will be presented by myself and others at the annual meetings of the Society for Risk Analysis and the Society for Medical Decision Making this fall. Future work will include the incorporation of environmental risks.

In the second project, Gary McClelland, Peter Ubel, David Asch, and I are attempting to assess the preferences of the public (and perhaps of patients or physicians) in the area of organ allocation. The current experiment is a nonrepresentative policy-capturing study in which participants give priority ratings to a number of hypothetical transplant candidates that vary in terms of eight attributes (e.g., time on waiting list, age, life expectancy with and without a transplant). Preliminary results will be reported in a paper at APPAM (Assoc. of Public Policy and Management) this fall. I have received a small faculty development grant to fund some of this work, and hope to use additional methods to study these issues in the future.

Finally, I am returning to some other environmental work that was on the back burner during my post-doc in medical decision making. Specifically, I am restarting a project involving the selection of habitat areas for protection-a topic of interest to various government agencies and private enterprises like the Nature Conservancy. I hope to have interesting results to report in the future.

Social Judgment Theory in Brazil

Claudia González-Vallejo, Ohio University

The most exciting Brunswikian experience of this past year was my trip to Brazil with Ken Hammond and Tom Stewart last summer. I was very fortunate to have been included in this event which entailed introducing and discussing Ken's ideas of his latest book on human judgment and uncertainty with policy makers associated/invited by FUNDAP, Sao Paolo. FUNDAP is a government institute that offers seminars and technical assistance to government institutions and conducts research in public policy. They have a vast number of interests and professionals working in areas such as public administration, agriculture, economics, education, energy, public finance, etc. We conducted a one-week seminar (I really mean one week, starting at 8:30 am and ending no earlier than 6 p.m., every day) where we introduced judgment and decision making research to the participants. We focused greatly on the main ideas of Ken's book regarding our actions and the duality of errors inherent in uncertain situations, coherence and correspondence theories of decision making, and judgment analysis. We used a wonderful program developed by Tom that takes the famous Taylor-Russell diagrams and shows, on line, how changing the uncertainty in the situation, moving the selection criterion, base rate, etc., affects the sizes of errors (false + and false -). This interactive Excel program also contains decision trees and 2 by 2 tables where one can add values and look at the problem from a decision analysis perspective. It is really nice!!

Brazil was crowded and with more problems than one can ever imagine (of course issues associated with poverty, but other problems as well regarding pollution, traffic, etc.). But the people were wonderfully warm and optimistic; the food was varied and delicious, and I would go back and do it all again if I had the chance!!!

My other research (some of it is more Brunswikian than others): Our paper on the management of ear infections, with Paul Sorum, Tom Stewart, John Chessare, and Jeryl Mumpower, was accepted for publication by Medical Decision Making. We are currently starting a second phase of that project involving physicians and colleagues in France-Gerard Chasseigne and Etienne Mullet. I have also continued to work on issues of choice behavior and the effect of the vagueness of information. Recent studies, some in collaboration with Dilip Soman and Sanjay Dhar, have been looking at the effect that vaguely stated discounts have on intentions to shop. I am also beginning to re-focus my energy on notions of judgment and calibration, some work with Josh Klayman and Jack Soll. In this area, I have recently taken a pretty Brunswikian view where probability judgments and confidence are based on a cue-utilization approach (see Koriat, 1997), and I think of them as two separate behaviors (not necessarily based on one underlying knowledge subjective variable, as many current models assume). I should have a draft soon which I would love to send out for comments to anybody interested.

Decision Making in Project Management

Kishore Sengupta
Naval Postgraduate School

In the past few years, my colleague Tarek Abdel-Hamid and I have conducted a number of studies that examine the micro-structure of decision making in software project management. We view decision making by managers in software projects as a case of dynamic decision making in the sense of Brehmer (1992) and Edwards (1962): The task requires several decisions made at different points over the life of the project; the decisions are interdependent; and the task environment changes autonomously, as well as in consequence of the subjects' decisions. Our research has, for the most part, focused on the issue of how managers make staffing decisions in software projects. The studies have been conducted primarily in laboratory settings, supplemented by field studies. The subjects participating in the experiments have ranged from graduate students with some experience in project management to software project managers with extensive experience in the domain.

The vehicle for conducting the studies is constituted by a computer-based microworld. The microworld embodies a systems dynamics model of the software development process, with a front-end "gaming" interface. The systems dynamics model was developed from field interviews with project managers in several software development organizations, supplemented by an extensive database of empirical findings from the literature. The model has been validated independently by five studies reported in the software engineering literature. The task environment represented by the model is complex (with more than 100 causal links) and dynamic. The environment changes autonomously as well as in consequence of a subject's actions.

In our experimental setup, a typical task runs as follows: A subject is told to manage the staffing level of a simulated software project from initiation to completion (the projects used in our studies are derived from actual software projects conducted at NASA and JPL). The progress of the project is tracked by using a set of reports delivered at intervals of every two calendar months (40 work days). In each period, the subject is required to decide on the staffing level for the next 40 days. The software then simulates the next 40 days, and provides updates on the status of the project. The subject makes another set of decisions for the next period, and so on, until the project is completed.

In selecting research questions of interest, we have adopted an approach similar to that of Morecroft (1988) wherein decisions in an organization are viewed through successive "filters," such as the individual's cognition, the organization's goal structure, the incentives provided to decision makers, etc. Some specific questions examined include issues such as the impact of feedback on decision making, the impact of unreliable information, and the effect of goals and incentive structures on the management of software projects.

The conclusions drawn from our research can be viewed from two perspectives: findings that are general to decision making by individuals in dynamic environments, and findings that are specific to the management of software projects. From the perspective of dynamic decision making, the following are some salient conclusions:

When operating in dynamic decision environments, individuals who receive outcome feedback perform poorly; they are out-performed by those who are aided by feedforward or cognitive feedback (Sengupta and Abdel-Hamid, 1993). In environments characterized by unreliable information, individual decision making is susceptible to self-fulfilling prophesies (Sengupta and Abdel-Hamid, 1996); subjects are also prone to a specific form of conservatism that we have labeled "conservative anchor-dragging" (Abdel-Hamid, Sengupta, and Ronan, 1993). Individuals have difficulty in coping with delays in the task environment. The extent of the difficulty depends on the degree to which the delay is visible, as well as the length of the delay (Sengupta, Abdel-Hamid, and Bosley, 1997).

From the perspective of software project management, some of the key conclusions are the following:

The goals of a project and the incentive structures offered to managers affect cost/schedule trade-off choices, as well as strategies for the acquisition and allocation of staff to different aspects of a project (Abdel-Hamid, Sengupta and Swett, 1997; Abdel-Hamid, Sengupta, and Hardebeck, 1994). These factors have a critical influence on the outcome of the project.

In our current efforts, we are examining the mental models of experienced project managers. We are especially interested in how they handle complex dynamic relationships in software projects, such as cost-schedule-quality trade-offs. Our preliminary finding is that while such individuals generally demonstrate a very good grasp of the development aspect of a software project, their understanding of the relationships underlying quality control, is much less sure (Sengupta, Abdel-Hamid and Swanson, 1997).

References

Abdel-Hamid, T., K. Sengupta, and D. Ronan. 1993. Software Project Control: An Experimental Investigation of Judgment under Fallible Information. *IEEE Transactions on Software Engineering*, **19**, 603-612.

Abdel-Hamid, T., K. Sengupta, and M. Hardebeck. 1994. The Impact of Reward Structures on Staff Allocations in a Multi-project Software Development Environment. *IEEE Transactions on Engineering Management*, **41**, 115-125.

Abdel-Hamid, T., K. Sengupta, and C. Swett. Goal Setting and Software Project Performance: An Empirical Investigation. Forthcoming, *MIS Quarterly*.

Brehmer, B. 1992. Dynamic Decision Making: The Control of Complex Systems. *Acta Psychologica*, **81**, 211-241.

Edwards, W. 1962. Dynamic Decision Theory and Probabilistic Information Processing. Human Factors, **4**, 59-73.

Sengupta, K., and T. Abdel-Hamid. 1993. Alternative Conceptions of Feedback in Dynamic Environments: An Experimental Investigation. *Management Science*, **39**, 411-428.

Sengupta, K., and T. Abdel-Hamid. 1996. The Impact of Unreliable Information on the Management of Software Projects: A Dynamic Decision Perspective. *IEEE Transactions on Systems, Man, and Cybernetics*, **26**, 177-189.

Sengupta, K., T. Abdel-Hamid, and M. Bosley. 1997. Coping with Staffing Delays in Software Project Management: An Experimental Investigation. Manuscript under review.

Sengupta, K., T. Abdel-Hamid, and D. Swanson. 1997. How do Experienced Managers Make Decisions in Software Projects? An Experimental Investigation. Manuscript under review.

Revenue Forecasting, Insanity, and Redistricting

Carmen Cirincione
University of Connecticut

I am working on a couple of research streams. Much of the work thus far focuses on the environment, but I have now begun to design, analyze, and/or write about the judgment and decision making side of the system.

Local Government Revenue Forecasting:

I have worked with two students, Gustavo Gurrieri and Bart Van de Sande, looking at the efficacy of various extrapolation methods in forecasting local government revenues. The analysis also examines the impact of length of series used to fit the model and the level of time aggregation. I have just begun writing a piece that utilizes the Integrated Contingency Model, which accounts for the properties of the task, the environment, and the forecaster, to explain the properties of local government officials' forecasts.

Insanity Defense:

For the last seven years, I studied the use, structure, and reform of the insanity defense. I am midway through a paper that uses a decision analytic framework and public misperceptions as to the structure of the defense and the flows into and out of the criminal justice/mental health systems to explain the design of insanity defense reforms enacted in the 1980s (40 of the 50 states and the federal government enacted reforms following during and/or immediately following the John Hinckly trial). I then provide evidence as to the efficacy of the enacted reforms in achieving the desired outcomes.

Congressional Redistricting:

Tom Darling, Tim O'Rourke, and I have built upon the work I presented at last year's Brunswik Society Meeting. Using a computer-intensive method, we have evaluated the bizarreness of congressional districting plans in five states: Alabama, Georgia, Mississippi, North Carolina, and South Carolina. A total of 100,000 computer generated congressional districting plans containing 820,000 congressional districts are used as a baseline for comparison in the assessment of adopted districting plans. We will soon sit down and design the judgment studies. Tentative title: Are Gerrymanders like Pornography: Do we know them when we see them?

Overconfidence Found in Estimates of Effort

Terence Connolly
University of Arizona

The only work we put out this year that looks recognizably Brunswikian in any substantial way was a paper with Doug Dean on software writers estimating how long it would take them to write their projects ("Decomposed vs. holistic estimates of effort required for software writing tasks," *Management Science*, 1997, **43**, 1029-1045). The thrust of this piece is a series of experiments in which programmers first estimated how long an upcoming task and its subelements would take to write, and then keeping good time records as they worked. They were, in general, close to unbiased (to our surprise, neither over-optimistic nor over-pessimistic), but they were hugely overconfident in their estimating skills, with much too narrow confidence intervals around their best guesses. Nearly half the outcomes were in the regions previously judged to have a 1% or smaller probability of occurring! This over-tightness appeared highly resistant to variations in question format and order, and to strenuous lecturing and feedback on the effect. It was, however, somewhat responsive to preliminary efforts to set plausible upper and lower limits of possible times. We discuss these findings in terms of both the practical (and high stakes) issue of estimating software tasks and the more theoretical matter of decomposng a natural task unit.

Judgmental
Accuracy and Task Predictability

Tom Stewart, University at Albany

With doctoral student Naiyi Hsiao, I have been working on an analysis of the relation between task predictability and accuracy of expert judgment. We now have data from 186 subjects in 18 studies of a wide range of experts including, among others, physicians, teachers, clinical psychologists, solar flare forecasters, and weather forecasters. Our measure of task predictability is R_e, the task multiple correlation. Admittedly, its an imperfect measure, since true task predictability can be lower, or even higher, than R_e, but we are prepared to defend it. Re for the tasks in the studies we have found ranges from .26 to .98.

For the data we have so far, the correlation between task predictability and accuracy is .87. Preliminary analyses indicate that for approximately 66% of subjects, accuracy is not significantly different from task predictability (accuracy is significantly lower than predictability for 25% of subjects and higher for 9%).

Our conclusion is that experts in a variety of fields operate near the limit of task predictability, and that task predictability explains about 75% of the variance in performance of experts in various fields. The data also show, as expected, that it is more difficult to reach the limit of task predictability when it is low than when it is high.

In other words, the major problem with expert judgment is with the environment, not with the experts.

The data we used are the same data that have been used for three decades to argue that expert judgment is poor because experts cannot outperform simple models. We argue that task predictability limits the performance of both the models and the experts The experts are doing about as well as they can, given the information that they have to work with.

10/3/97