

THE BRUNSWIK SOCIETY

Newsletter 1998

Volume 13 Number 1

Albany, New York

Fall 1998

This edition of the Brunswik Society Newsletter was edited by Tom Stewart (t.stewart@albany.edu) with the editorial assistance of Sue Wissel (SW831@CNSIBM.ALBANY.EDU) and supported by the Center for Policy Research (University at Albany, State University of New York). This electronic version was prepared by Junseop Shim. This newsletter was reformatted as a pdf file in May 2022. Some minor corrections were made.

Contents

Integration of Brunswik and Gibson

Kim Vicente Toronto, Ontario, Canada

Perceptions of Interest: A Lens Model Analysis

James Athanasou Sydney, Australia

Repeated Judgments of Interest in Vocational Education: A Lens Model Analysis

James Athanasou Sydney, Australia

Individual Quality of Life (IQoL)

Dick Joyce Allschwil, Switzerland

Team Adaptation to Time Pressure in Dynamic Environments

Leonard Adelman Fairfax, Virginia

Egon Brunswik and Edward Tolman

Nancy Innis London, Ontario, Canada

Accuracy in personality judgment

David Funder Riverside, California

Customer-Salesperson Relationships and Teaching the Lens Model

Dale Rude Houston, Texas

Medical decision making and judgment

Robert M. Hamm Houston, Texas

Why Do Doctors Disagree When the Evidence is Clear? - Physicians' Judgments of the Outcomes of Therapy for Patients with Heart Failure

Roy M. Poses, Providence, RI Maria Woloshynowych, London, United Kingdom D. Mark Chaput de Saintonge, London, United Kingdom

Nomothetic judgment analysis

Mandeep Dhami London, United Kingdom

Brunswikian research at the University of Connecticut

Jim Holzworth Storrs, Connecticut

Bowling Green Pastures

Michael Doherty Bowling Green, Ohio

Research in Uppsala

Peter Juslin Uppsala, Sweden

Design and training problems in dynamic, high-technology and high-consequence environments

Alex Kirlik Atlanta, Georgia

Lens model research at Buffalo

Ann Bisantz Amherst, New York

Latest News on Brunswikian Music Psychology

Patrik Juslin Uppsala, Sweden

Judgment research at Vanderbilt

Jim Hogge Nashville, Tennessee

Performance of Physicians at Judging Survival in Congestive Heart Failure

Wally R. Smith Richmond, Virginia

Research at University College London

Clare Harries London, United Kingdom

Brunswik and Complexity

Ray W. Cooksey Armidale, NSW, Australia

"The Essential Brunswik: Beginnings, Explications, Applications"

Ken Hammond, Boulder, Colorado, Tom Stewart, Albany, NY

Judgments under stress

Kenneth Hammond Boulder, Colorado

Policy Capturing and Multiple Regression: A Married Couple?

Ulrich Hoffrage Berlin, Germany

Overconfidence, causal systems, and other people's feelings

Joshua Klayman Chicago, Illinois

Simulation, stress, and visualization research in Georgia

Rob Mahan Athens, Georgia

Learning and performance in interactive dynamic decision tasks

Bud Gibson Ann Arbor, Michigan

Accuracy of physician judgment

Neal Dawson Cleveland, Ohio

Current activities of a more or less Brunswikian nature

Alexander J. Wearing Melbourne, Australia

Integration of Brunswik and Gibson

Kim Vicente

Toronto, Ontario, Canada

The primary research activity I have been involved in during the past year (with valuable and significant input from Tom Stewart) is the integration of Brunswikian and Gibsonian approaches to ecological psychology. This work has led to a paper currently under review that is entitled, "Building An Ecological Foundation for Experimental Psychology: Beyond the Lens Model and Direct Perception." An abstract follows:

Cumulative theory in psychology has been hard to come by. In this article, a unified, ecological foundation for experimental psychology is proposed by integrating Brunswikian and Gibsonian theories. By identifying the constraints imposed by the environment on cognitive processes, this framework provides a single map for situating and relating research from distinct areas, including: automaticity, direct perception, expertise in memory recall, judgment and decision making expertise, judgment under incomplete knowledge, perceptual learning, and problem solving. This map consists of two axes, one defining the degree of goal-relevant distal structure in the environment, and another defining the degree of proximal stimulation that is available to specify that structure. These dimensions specify feasible limits on performance and on the strength of the coupling to the environment, respectively. The conceptual and empirical support for the framework is reviewed, and its benefits are described.

Perceptions of Interest: A Lens Model Analysis

James Athanasou

Sydney, Australia

This study tested recent German theories of the nature of human interest, in which it is hypothesized that individual interest is composed of cognitive, emotional and value components. Using an idiographic design based on representative sampling of a classroom ecology, 10 judges rated 108 student profiles for the level of actualized interest. The profiles were obtained from 27 random experience samples using 17 cues: level of skill, knowledge, success, familiarity, confidence, concentration, understanding, satisfaction, happiness, excitement, effort, enthusiasm, enjoyment, desire, determination, importance, and extent of freedom. The 27 profiles were presented in four blocks and judges were reasonably reliable in their 27 judgments with a median inter-trial correlation of 0.83 and a coefficient alpha for the 17 ratings of 0.95. A lens model analysis was used to decompose judgments across repeated situations in order to determine the key components of actualized interest. Lens model parameters, such as R-squared, ranged from 0.94 to 0.55, and cognitive consistency ranged from 0.96 to 0.74. Based on the relative beta weights, the most important indicators of interest were ratings of effort, happiness, desire, familiarity, enthusiasm, importance and enjoyment. Results supported the emotionality and value components of actualized interest, but not the knowledge emphasis in German theories.

Repeated Judgments of Interest in Vocational Education: A Lens Model Analysis

James Athanasou
Sydney, Australia

The purpose of this study was to decompose student decision-making about the levels of subject interest in vocational education. Ten technical and further education students made 120 judgments of the level of classroom interest from actual protocols of responses. These paid participants repeated the judgments after receiving details of their personal judgment policy in graphical form and the actual environmental relationships in graphical form. Judges were operating in a relatively predictable environment (Multiple R=0.795). Cues involved the quality of teaching, the importance of the subject, ability, difficulty of the subject, whether the course was liked, study time and homework time. Results were analyzed in terms of a lens model in which judgment achievement is a function of the task properties, cognitive control and knowledge. Students' mean level of achievement correlation (Fisher Zr) increased from 0.31 to 0.39. There were no significant differences in the levels of lens model indices (G, Rs, C) from pre- to post-information judgments. Results suggested that students overcompensated in their efforts to maximize judgment accuracy and that they were not able to make full use of the entire range of cues. The results have implications for students' perceptions of their interest in vocational education subjects.

Individual Quality of Life (IQoL)

Dick Joyce
Allschwil, Switzerland

My colleagues at the Royal College of Surgeons in Ireland, Hannah McGee and Ciaran O'Boyle, and I have reason to believe that the edited book with this title will appear in time for the Fall round of relevant meetings. Its contents describe methods and observations with several new instruments designed to get inside the individual black box that is primarily relevant to descriptions of IQoL. The Schedule for the Evaluation of IQoL (SEIQoL), one of the first of these, is an application of SJT. It is on the short list of instruments recommended by WHO and is believed to be in use or to have been used in something like 200 studies, some of them in parts of the world where QoL might be considered poor if not totally absent (like body temperature, everyone has a QoL). I have also been applying SJT to the study of non-specific factors in therapeutic interventions. Summarizing preliminary findings in a paper delivered at the monastery of Einsiedeln, entitled "Is God a Placebo?", the slide projector - perhaps not surprisingly - declined to cooperate.

Team Adaptation to Time Pressure in Dynamic Environments

Leonard Adelman
Fairfax, Virginia

My students (DeVere Henderson and Sheryl Miller) and I have been examining how teams adapt to increasing levels of time pressure in a study we conducted this year. Conceptually, we are using the multi-level, lens model that Brehmer and Hagafors developed in 1986 to study staff decision making, and that Rob Mahan and his students at the University of Georgia have used in the research they've described in the last two conferences. Seven, three-person teams participated in our study. Each team was composed of ROTC cadets, who participated in the study for two hours per week for seven weeks. Our task was a dynamic, aircraft identification

task. Two staff members (and a leader) had to track aircraft on their screens, pass information about the aircraft to each other, and make recommendations about the aircraft's level of hostility, which the leader could then use to make judgments while the aircraft were on the screen.

Consistent with Ken Hammond's recent presentations, we hypothesized that increased time pressure (i.e., less time to make a judgment about each aircraft), would (1) affect teams' correspondence constancy (i.e., performance would deteriorate), but that (2) teams would adapt (perhaps in different ways) in an effort to maintain it. That is exactly what we found. Performance decreased, although not as quickly or precipitously as predicted. In addition, there were few significant differences in the teams' overall performance scores. Teams did, however, adapt (or not) in different ways to increased time pressure. Three of the seven teams tried to continue performing the task as trained regardless of the time pressure; that is, the subordinates kept sending identification recommendations to the leader for all aircraft. (We think these teams tried to work faster and use simpler, more intuitive organizing principles, but we're still doing analyses.) In contrast, two teams simplified the task by having each subordinates make recommendations for only half the aircraft. And in two teams, the leader took over the entire decision making task by having subordinates only send information about the aircraft, not recommendations.

In addition, the leaders made clear speed-accuracy trade-offs in an effort to maintain performance. For example, in the condition with the greatest time pressure, the leader of one of the two leader-controlled teams made judgments for more aircraft than any other team, but had the lowest achievement score (ra). In contrast, the leader for the other leader-controlled team had the highest achievement score, but made the fewest number of judgments. Utilization of these (and other) adaptation strategies resulted in essentially equivalent levels of performance overall because none of the teams were able to maintain both speed and accuracy under high time pressure. More generally, we consider the study as just a first step toward applying the concepts of vicarious mediation and vicarious functioning to (a) understanding how teams adapt in efforts to maintain correspondence constancy, and (b) designing systems to support them.

Egon Brunswik and Edward Tolman

Nancy Innis
London, Ontario, Canada

In 1933-4 while on a sabbatical leave in Europe, Edward C. Tolman (1886-1959), the eminent American learning theorist, spent several months in Vienna. As part of the research for a biography of Tolman I'm working on, I am examining the ideas of people associated with the Vienna Circle and the Institute of Psychology at the University of Vienna at the time Tolman was there.

To put Tolman in context for those who aren't familiar with his work, following an MIT degree in engineering, he received a Ph.D. in psychology from Harvard, under the supervision of Hugo Munsterberg, for studies of memory. In 1918, he joined the faculty at the University of California, Berkeley, where he began to carry out research with animal subjects and to develop his own theory of learning - purposive behaviorism. He presented his system in detail in his 1932 book, *Purposive Behavior in Animals and Men*. The book was highly acclaimed, and Tolman's ideas were receiving widespread attention when he began his sabbatical in 1933.

In Vienna, Tolman spent much of his time with the group at the Institute of Psychology headed by Karl Buhler and soon developed a close professional association with Egon Brunswik. The two men met frequently in Viennese cafes to discuss psychology, and it soon became evident that Brunswik's probabilistic functionalism was very similar to Tolman's behavioristic approach. They began collaboration on a paper identifying the common features of their ideas, and "The organism and the causal texture of the environment" was published in

the Psychological Review in 1935. The suggestion that the relationship between cues and objects or signs and goals was probabilistic rather than one-to-one led to major changes in animal learning research. The most positive outcome of Tolman's sabbatical in Vienna was that Egon Brunswik eventually obtained a position in America. As a result of his association with Tolman, Brunswik received a Rockefeller Foundation fellowship and spent 1934-35 at Berkeley. By the mid-1930s, political conditions in Europe were worsening and prospects for advancement there were slim. Tolman began a campaign to find a position for Brunswik, and the Berkeley department was able to offer him a faculty appointment which he took up in 1937.

My current research involving Brunswik and Tolman focuses on examining the similarities and differences in their theoretical positions (as outlined in the Psychological Review paper) during the 1930s, and the influence each man had on the work of the other when they were colleagues at Berkeley.

Accuracy in personality judgment

David Funder
Riverside, California

I am putting the final touches on a book about my approach to the study of accuracy in personality judgment, to be published by Academic Press. It presents the "Realistic Accuracy Model" and its mostly-Brunswikian philosophical underpinnings. If anyone who reads this would like one or more chapters to read in draft form I would be happy to e-mail them. However, the cost for this is that one must promise to provide criticism in time for me to do something about it. The tentative chapter titles are:

1. Approaching accuracy (general introduction)
2. The very existence of personality
3. Error and accuracy in the study of personality judgment
4. Methodological and philosophical considerations
5. The process of accurate personality judgment
6. Moderators of accuracy
7. Self-knowledge
8. Prospects for improving accuracy

Customer-Salesperson Relationships and Teaching the Lens Model

Dale Rude
Houston, Texas

I have been engaged in several activities that may be of interest to the Society. The first is a research project with Eli Jones, a marketing faculty member whose primary research area is relationship selling. We are doing a

dyadic study to see if customer-salesperson relationship variables (e.g., similarity, frequency of contact, trustworthiness, length of relationship) predict how well the salesperson understands the customer's decision strategy for making a purchase. Customer and salesperson rate the same series of policy capturing scenarios, the salesperson attempting to predict the customer's responses. The dependent variable is the agreement between salesperson and customer's policy/decision strategies.

The second activity involves software and text materials that I have written for teaching Brunswik's lens model to undergraduate business majors. These students have a limited understanding of statistics and no research methods background. One software program is an APL DOS-based horse race task (modeled after the Harmon and Rohrbaugh horse race task) which I use for introducing the lens model. Students make bets and judgments of order of finish and odds using four cues including speed rating, jockey winning percentage, and post position. Their judgments are then analyzed to produce achievement indices, etc. Students also compare their performances to that of an expert handicapper. The second software program is APL DOS-based and is used in my classes for a team project. Teams of three students identify a decision environment and problem for study and collect judgments, cues, and outcomes. After the data are entered, the software performs the statistical analyses and produces the relevant lens model indices. The students then write a 20-page report summarizing their findings and do a class presentation. When finished, the students thoroughly understand the lens model.

The text materials consist of class notes, assignments, and problems that I use. These include a review of relevant statistical concepts (mean, standard deviation, correlation, regression, etc.), the lens model, and some application problems. I have simplified the lens model terminology because students find some of the traditional lens model terminology to be very confusing (e.g., criterion is commonly confused with cue). For example, I have substituted "outcome" for "criterion," "cue validity" for "ecological validity," and "outcome predictability" for "environmental predictability."

Anyone who would like the software and/or teaching materials should email me and provide a "snail mail" address. I will send you several floppy disks containing the software and teaching materials.

Medical decision making and judgment

Robert M. Hamm
Houston, Texas

This year I have worked on projects that address judgment and decision making in the medical context.

Patients' probabilistic inference. In this study (Hamm, R.M., and Smith, S.L. (1998). The accuracy of patients' judgments of disease probability and test sensitivity and specificity. *Journal of Family Practice*, 47, 44-52), patients read a vignette of a person seeing a doctor with a given complaint. The major disease that could produce that complaint, and the test typically used to see if the patient has that disease, were described. Patient was asked to estimate prior probability $[p(\text{disease})]$, test sensitivity $[p(T+|D)]$, test specificity $[p(T-|D-)]$, and post-test probability $[p(D|T+)]$. Knowledge (prior probability and test characteristics) was inaccurate; probabilistic inference (using patient's own judgments) was inaccurate; and past experience with the disease improved the accuracy only slightly. We argued that this demonstrates a need to educate patients explicitly about the possibility of inaccurate test results.

Improving physician judgment. We reviewed past attempts to improve physician decision making through the applications of judgment research or decision analysis (Hamm, R.M., Scheid, D.C., Smith, W.R., and Tape, T.G. (in press, 1998). Opportunities for applying psychological theory to improve medical decision making:

Two case histories. In G.B. Chapman and F. Sonnenberg (Eds.), *Decision Making in Health Care: Theory, Psychology, and Applications* (pp ?-?). New York: Cambridge University Press), and focussed in particular on two projects. One of them gave physicians cognitive feedback to train them to make more accurate probability judgments, but did not change the rate of a criticizable action (prescribing antibiotics for sore throats that are probably due to viral infections). (This result was similar to one reported to this group last year by Tom Taylor.) The other study we discussed gave physicians an accurate (state of the art!) estimate of the probability a patient would die of their illness, and some info about patient values, but it did not reduce the amount of "futile" end of life interventions. Our conclusion might be characterized as "it didn't prove that our science is useless, we just have to try harder,' or 'they did not analyze the situation well enough to discover what was really going on so the information/training provided was irrelevant." What, then, would be relevant? We hope that our paper, which will appear in a volume sponsored by the JDM society, will encourage researchers to look anew at the hard problems encountered when trying to use our best scientific understanding to improve applied decision making.

Analyzing a medical decision. We are analyzing a decision point that occurs when physicians screen women for precursors of cervical cancer (Hamm, R.M., Loemker, V., Reilly, K., Johnson, G., Dubois, P., Staveley-O'Carroll, K., Brand, J., Owens, T., and Smith, K. (in press, September, 1998). A clinical decision analysis of cryotherapy versus expectant management for cervical dysplasia (CIN 1). *Journal of Family Practice*). If a particular condition is found, one could treat now or wait and see if it goes away and treat only if it doesn't.. If we had confidence the patient would return for the required follow up, then waiting would be better. We are doing a study of 300 patients, to see if it is possible to predict who will faithfully return for follow up in the coming year. In addition to data from the medical record and from a questionnaire the patient fills out, we have asked the physician and nurse to judge the likelihood the patient will return.. We will analyze the accuracy of these experts' judgments, as well as produce an environmental model. Is there enough info available to the physician to afford a prediction that could make a difference in who gets treated?

Why Do Doctors Disagree When the Evidence is Clear? - Physicians' Judgments of the Outcomes of Therapy for Patients with Heart Failure

Roy M. Poses, Providence, RI

Maria Woloshnowych, London, United Kingdom

D. Mark Chaput de Saintonge, London, United Kingdom

Physicians vary in their use of treatments and frequently fail to use specific treatments despite strong evidence supporting their use. For example, several studies have shown that physicians do not prescribe angiotensin converting enzyme inhibitor (ACEI) drugs for many patients with chronic congestive heart failure (CHF) and systolic dysfunction, a clinical syndrome characterized by the inability of the heart to pump blood adequately for the body's needs, despite data from several large controlled trials that ACEI's prolong survival and decrease morbidity for such patients while causing few adverse reactions. The reasons physicians fail to do what they apparently ought to do in this situation remain unclear. We postulated that physicians may base treatment decisions on their judgments of the probability of relevant outcomes conditional on whether or not the treatment were to be given, and that problems they have judging these probabilities may partially explain practice variation and failure to use treatments supported by evidence. For example, physicians may base decisions to prescribe ACEI's on the difference between their judgments of the probability of survival were ACEI's given and those of survival were ACEI's withheld (the difference is the judged survival advantage due to ACEI's), and on their judgments of the probability of adverse drug reactions (ADR's) were the drugs to be given. Further,

physicians' judgments of these outcome probabilities for particular patients may not be based on clinical factors that evidence suggests predicts these rates, and also may be based on factors that might plausibly appear to be predictive, but in fact are not. The purpose of the current study was therefore to assess physicians' judgments and decisions about ACEI use for patients with CHF and systolic dysfunction.

Our physician subjects were 20 general practitioners, 21 hospital-based internists other than cardiologists, and 19 hospital-based cardiologists practicing in London, England, selected essentially randomly. We used a standardized written instrument to assess: the physicians' general propensity to treat patients with moderate or severe CHF with systolic dysfunction (measured as the proportion of 100 such patients they would treat); and their judgments of the overall rates of survival were these patients to be treated or not treated with ACEI's, and of the overall rates of ADR's and serious or fatal ADR's were these patients to be treated with ACEI's. We then asked the physicians to consider 27 case vignettes. The vignettes were constructed such that selected clinical characteristics systematically varied across them. These included characteristics that evidence suggests predict either good or bad outcomes for CHF patients conditional on treatment (e.g., class IV CHF predicts a greater survival advantage due to ACEI treatment, while hypotension predicts a higher likelihood of ADR), and characteristics which do not predict such outcomes (e.g., a history of a kidney stone or of gastroesophageal reflux, or low but normal blood pressure). These latter variables were chosen to represent mild co-morbidity which should not affect outcomes for CHF treated by ACEI's, or because they somehow resembled factors which were predictive, but were not themselves predictive.

We found that doctors' overall propensity to treat with ACEI's was very high and showed little variability across physicians. However, there was little agreement between individual physician's stated propensity to treat patients with either mild-to-moderate or severe CHF and the rate he or she treated corresponding case vignettes. For the majority of physicians, we found relationships both between judged survival advantage due to ACEI's and ACEI decisions for the case vignettes, and between judgments of ADR probability and these treatment decisions.

How physicians weighted clinical variables when making judgments of outcome rates varied widely. The majority of physicians failed to use the severity of heart failure (in terms of the patient's functional status) as a predictor of survival advantage despite strong evidence that patients with the worst functional status have the greatest survival advantage. Some physicians thought mild, irrelevant co-morbidities like a history of a kidney or gastro-esophageal reflux predicted ADR's. Twenty percent (12/60) of physicians used at least one positively predictive or non-predictive variable as a significant negative predictor of survival advantage, and the same proportion of physicians used at least one non-predictive variable as a significant positive predictor of ADR risk.

In summary, asking physicians about their general propensity to treat may reflect an ideal but not their actual treatment decisions. Judgments of relevant outcome rates may relate to and perhaps determine how doctors make decisions in some cases. When judging outcome rates, physicians may fail to take into account predictive variables and/or take into account non-predictive variables. Appreciating how doctors make judgments and decisions may lead to better interventions to decrease unnecessary practice variation and improve quality.

Nomothetic judgment analysis

Mandeep Dhani
London, United Kingdom

Judgment Analysis (JA) within the framework of Social Judgment Theory is commonly (traditionally) conducted at the idiographic level, using linear regression modelling techniques. Over the past year, in an attempt to apply JA to expert decision making in the legal domain, I stumbled upon many practical problems concerning collection and analysis of judgment data.

The first study investigated British magistrates' bail decision making on hypothetical cases, using an orthogonal cue set. The high case to cue ratio necessary for regression modelling meant that I had to either reduce the number of cues studied or increase the number of cases. Magistrates were unwilling to participate in either a time consuming task, or an unrealistic task. The second study investigated magistrates' bail decision making in the courtroom, and used a fully representative design. Here, there were high inter-cue correlations between some cues; and in the courtroom, data was missing on some legally important cues, which created problems for conducting regression modeling. The cues with a lot of missing data had to be omitted, and it was difficult to ascertain which cues were used.

After brief consideration of abandoning JA and/or legal decision making, I decided instead to test alternative ways of conducting JA.

Firstly, rather than conducting JA at the idiographic level, I developed and tested JA at the nomothetic level. Individual magistrates made judgments on a smaller case to cue ratio, namely 3:1. The judgments of the whole group were then analyzed using regression modeling techniques, and the model of the group was cross-validated at the idiographic level on a set of holdout cases. The cross-validation procedure revealed that the model of the group adequately represented the policies of individual magistrates, taking into consideration each magistrate's level of consistency; and it also identified a small sample of magistrates for whom idiographic level analysis may be necessary. The ability of the nomothetic level analysis to adequately represent individual magistrates' policies is not surprising as previous studies using JA at the idiographic level have found that clusters of judges with similar policies emerge.

Secondly, rather than using regression modeling techniques to model judgment, I adapted and tested a simple heuristic referred to as Take The Best (TTB) which is one of a family of algorithms developed within the framework of Probabilistic Mental Models, by Gigerenzer & Goldstein (1996). Regular readers of the Brunswik Newsletter may recall that Ulrich Hoffrage mentioned such models in the 1996 issue, and attendants at the 1997 Brunswik Meeting in Philadelphia may recall hearing Gerd Gigerenzer speak on this topic. The TTB I adapted and tested uses judgment data in a frequency format and can cope with inter-cue correlations and cues with missing data. The TTB model cross-validated well on a set of holdout cases. The representation of human judgment provided by the TTB model is one of limited information search and one reason decision making, and so suggests that magistrates are non-additive, non-linear and non-compensatory. Given the limited demand on information processing capacity and complexity, the TTB model provides a psychologically more plausible account of human judgment.

So, over the past year, I've used JA, but not as we know it. In the year to come, I aim to continue using simple heuristics in modeling human judgment.

Anyone interested in copies of the above studies may contact me, and I would welcome any comments.

Brunswikian research at the University of Connecticut

Jim Holzworth
Storrs, Connecticut

Research in the Brunswikian tradition continues at the University of Connecticut. My colleagues Steven Mellor and Jim Conway and I have just completed data collection in a judgment study investigating people's inclinations to be represented by labor unions. Employed persons (not union members) were asked to make judgments concerning how likely they would be to vote in favor of (or against) union representation. Data analyses will test a weighted decision model based on the relative importance of costs and benefits of union representation. Design details of this study were presented in last year's newsletter. In another union-related judgment study, Mellor, Dan O'Shea and I have begun investigating judgments concerning crossing a picket line. We are in the design phase of this study.

Studies testing premises of Cognitive Continuum Theory (CCT) are in various stages. Some of the design details were mentioned in last year's newsletter. These studies 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. Within-subject designs are employed. In an aesthetic judgment study, 24 participants viewed different styles of art (representational and nonrepresentational), "thinking aloud" while viewing each painting. After rating each painting, participants were asked to justify their responses. Julia Pavone (fine arts) is assisting me with this project. We are in the data analysis stage. Along with Janet Barnes-Farrell, I am continuing my research on worker performance appraisal. Our study has participants viewing work samples of restaurant waitresses presented in several ways (videotapes, written transcripts, and summary data). Each participant evaluates overall performance of the waitress and gives oral justification. Data collection is under way. A chromosome classification study is finally getting off the ground. My colleague Judy Brown and I have 200 photo images of human chromosomes (all 23 pairs), 100 of each sex. Study participants will be asked to sort some number of these images into two sets (male/female) based on presence of a Y chromosome. Participants will "think aloud" while sorting. In each experiment, verbal protocol analysis of "think aloud," justification, and evaluation data will be done to test premises of CCT.

Tom Stewart and I are looking for judgment data sets which include criterion information for further testing of my smart ridge regression technique (combining human judgment with ridge regression; OBHDP, December, 1996). Anyone willing to share data may be richly rewarded. I hope to have an opportunity to discuss some of my research at our Brunswick Meeting in November.

Bowling Green Pastures

Michael Doherty
Bowling Green, Ohio

June 1 brought my formal retirement from Bowling Green.

In the past year, a student, Greg Brake, completed his PhD dissertation, in which Brunswikian principles were brought to bear on calibration research. His results were consistent with the easy-hard effect, but the lens model approach allowed an independent specification of task difficulty. This is, to my knowledge, the first time that difficulty has been specified independently of percent correct. Part of Greg's dissertation research used a baseball domain, as did the work he presented last year, and found elegant underconfidence on games that were

relatively predictable, and overconfidence (not as cleanly) on the harder games. He also used the prediction of roommates in an additional calibration task. In that study, the students showed bias (most of the data curve under the leading diagonal) Difficulty in both domains was operationalized as Re.

A paper that Ken Shemberg and I gave as a poster at JDM in Chicago was accepted for publication in the Journal of Clinical Psychology. In that paper we tried a framing manipulation to bias diagnostic judgments of 136 practicing clinicians. The Brunswikian aspect of the paper was the care given to object sampling, in that we used 9 different target descriptions. The clinicians were affected by the biasing question format, but in a pretty sensible way; they were much less likely to assent to the extreme diagnosis of psychopathology than the less extreme (more common) "psychologically unhealthy."

We (Greg Brake and Dave Slegers) are continuing the effort to generalize Gigerenzer's fast and frugal model. We are scheduled to present this work at Psychonomics.

We (Jack Mynatt and I) finally finished an Intro book, after many years. It will dismay Ken and other members of this society to see us pay so much attention to classical design and single variable studies, but at least we have included a critique of classical design from an explicitly Brunswikian perspective. In Ken's marvelous 1966 opening chapter, he indicted Intro books as one of the reasons that representative design has not had a higher profile, but I do not see how one could present the current state of psychology for INTRODUCTORY students right out of high school from a Brunswikian point of view. Perhaps a younger member of the society might take on such a task with a much fresher perspective.

Research in Uppsala

Peter Juslin
Uppsala, Sweden

Our current research in Uppsala can be summarized in four points: First, we follow up on previous research. Here are two examples. We (Anders Winman, Henrik Olsson and I) have recently finished a first version of a paper that contains a "meta-analysis" of all published or otherwise available data on realism of confidence in studies with two-alternative general knowledge items. The study investigates a) if representative vs selected item samples has an effect on over/underconfidence, b) if there is evidence for a cognitive overconfidence bias when the items have been representatively selected, and c) if there remains a substantial hard-easy effect in these data once that we control for selection effects and more trivial statistical effects. We (mainly Anders Winman and I) are also starting to analyze a data set that provides further comparisons between Thurstonian and Brunswikian error in judgment, e.g., with respect to additivity, over/underconfidence in interval estimation, and "ambiguity avoidance". Henrik Olsson is continuing the work on the sensory sampling model, among other things, applying it to the case of perceptual bias.

Second, we (mainly Anders Winman and I) have performed two experiments that aim to investigate the nature of the cognitive representations that underlie judgment in MCPL. We are specifically interested in finding ways to figure out when judgments are mediated by rule-based processing vs memory-based processing. This analysis is still in progress.

Third, we (Magnus Persson and I) are working on an exemplar-based model of subjective probability assessment, entitled PROBEX. The model is intended to describe memory-based judgments and predicts decisions, point-estimates, subjective probabilities, and response times. We are currently finishing the computer-simulations and the data collection. Again, this is work on the way right now.

Finally, we (mainly Pia Wennerholm and I) are performing studies on base-rate use in judgment, both in more simple categorization designs and more ecologically relevant contexts.

As is evident much of this research is currently in progress, so the fertility of many of our efforts is still uncertain. This is only as it should be, of course."

Design and training problems in dynamic, high-technology and high-consequence environments

Alex Kirlik
Atlanta, Georgia

We continue our studies of design and training problems in dynamic, high-technology and high-consequence environments. Working with the U.S. Naval Training Systems Division, we have investigated the factors limiting judgment performance in target identification tasks, such as those performed by Navy AEGIS operators. We have found that high and low performing participants in such tasks are distinguished primarily by execution consistency, rather than by differences in task knowledge. Our experimental work in this domain has also demonstrated the potential of part-task training and automated feedback technology for augmenting the "over the shoulder" coaching predominantly used in operational training. These studies appear in Kirlik, Fisk, Walker and Rothrock (in press) and in Bisantz, Kirlik, Gay, Phipps, Walker and Fisk (1998). Also in this context, Ann Bisantz (bisantz@eng.buffalo.edu) has completed a dissertation investigating how experience in making judgments in a dynamic, interactive task influences subsequent performance on evidential reasoning tasks within the same context. Also in a naval context, Richard Strauss (strauss@isye.gatech.edu) and I are working with the Johns Hopkins Applied Physics Lab for ONR on situation awareness issues in submarine operations. Strauss has performed task analyses and has constructed a dynamic laboratory simulation of the task of coming to periscope depth. We will soon begin experimentation to understand how operators cope with uncertainty in this task environment, in order to develop new concepts for interface design.

We have also been working with NASA Ames on human-automation interaction issues and training problems associated with the planned transition to a "free-flight" environment. The control systems of modern "glass cockpit" aircraft have such complexity that it is often said that pilots frequently ask themselves these questions: "What is it doing?," "Why is it doing that?" and "What is it going to do?" From a Brunswikian perspective it can seem as if many high-technology systems have been structured specifically to defeat the judgment strategies and heuristics evolution adopted for meeting the demands of the natural world.

Based on the recent dissertation by Degani (degani@mail.arc.nasa.gov), we have proposed a modeling technique that can be used to document the complex structure of mode-based control systems in order to identify any interface design features that may contribute to mode confusion and mode error (Degani, Shafto, and Kirlik, in press). It may be interesting to note that this publication in an aviation psychology journal consists solely of an environmental modeling technique.

We have recently begun work with NASA Ames on flight crew training problems likely to arise as a result of the free-flight concept currently envisioned for U.S. aviation operations. Free-flight is a new traffic control regime which would give individual flight crews much more control over their flight trajectories than is currently allowed. In the current centralized control system, air traffic control (ATC) directs aircraft along a limited set of trajectories or "highways in the sky," and any requested deviation from the planned trajectory (e.g., to fly a more direct path or a more fuel-efficient altitude) must be first approved by ATC. One interesting aspect of free-flight is that individual flight crews will be given the opportunity to detect and resolve potential

conflicts without ATC intervention, using cockpit displays of traffic information and a set of negotiation rules (e.g., the aircraft on the right has the right of way). Currently we are examining how judgment analysis might provide resources for investigating training issues arising from free-flight.

Finally, Ann Bisantz and I have written a chapter providing an overview of how cognitive engineering has contributed to our understanding of cognition by examining experiential and environmental aspects of adaptation.

References:

Bisantz, A.M., Kirlik, A., Gay, P., Phipps, D.A., Walker, N., and Fisk, A.D. (1998). Modeling and analysis of a dynamic judgment task using a lens model approach. Ms. currently in review to IEEE Systems, Man, and Cybernetics.

Bisantz, A.M. (1998). Modeling environmental uncertainty to understand and support dynamic decision making. Unpublished doctoral dissertation, School of Industrial & Systems Engineering, Georgia Institute of Technology.

Degani, A., Shafto, M., and Kirlik, A. (in press). Modes in human-machine systems: review, classification, and application. *International Journal of Aviation Psychology*.

Degani, A. (1997). Modeling human-machine systems: On modes, error, and patterns of interaction. Unpublished doctoral dissertation, School of Industrial & Systems Engineering, Georgia Institute of Technology.

Kirlik, A., Fisk, A.D., Walker, N. and Rothrock, L. (in press). Feedback augmentation and part-task practice in training dynamic decision making skills. In J. Cannon-Bowers and E. Salas (Eds.), *Decision Making in Complex Environments*. APA Press.

Kirlik, A. and Bisantz, A. (in press). Cognition in human-machine systems: Experiential and environmental aspects of adaptation. In P.A. Hancock (Ed.), *Handbook of Perception & Cognition Vol. 17: Human Performance & Ergonomics*. Academic Press.

Lens model research at Buffalo

Ann Bisantz
Amherst, New York

In the past year, I have been involved with several judgement and decision-making research projects. While at Georgia Tech, I worked with Alex Kirlik on research which investigated how performance on several different judgment tasks might be adaptive to the nature of the uncertainty in the environment when people are exposed to that uncertainty through explicit experience acting in the environment, or by task formats which allow people to tap into their experiential knowledge of uncertainty. This work was based on prior research (e.g., Cosmides and Tooby, 1996; Kirby, 1994; Klayman and Ha, 1987) which suggested that performance on tasks such as evidential reasoning, and rule verification or hypothesis testing might reflect adaptation to environmental probabilities rather than non-normative biases, or might approach normative solutions when task formats mimic people's natural experience with uncertainty. The research showed some adaptivity to environmental probabilities in a categorization task and an evidential reasoning task. Other research with Kirlik, Neff Walker,

Dan Fisk, Donita Phipps, and Paul Gay at Georgia Tech used the Lens Model to model judgments in a complex Naval Command and Control Task. This research was interesting because, due to the dynamic nature of the environment, and the degree to which participants could choose when they made judgments (in this case, identifying aircraft), and what information they had available to make a judgment, the cues and cue values for a given judgment were not consistent across judges. This necessitated the use of a different environmental model, and thus different values of Re , for each judge. Through the Lens Model analysis we found that differences in performance between good and poor performers were attributable to differences in consistency rather than differences in knowledge; this was consistent with results from an analysis of individual errors.

Finally, at SUNY Buffalo, I am working with two graduate students who are utilizing Lens Models in their research. Gordon Gattie has begun working in collaboration with faculty at the SUNY at Buffalo Dental School to develop a computer based training tool for dental students to help them learn to diagnose different oral diseases based on clinical photos, and plans to incorporate and test aspects of cognitive feedback in this application. In the context of a project through the Center for Multi-source Information Fusion at SUNY Buffalo, Younho Seong is interested in modeling human trust in automated systems that may be degraded or sabotaged. We are exploring the use of the Lens Model to capture aspects of calibration between the extent to which a human operator relies on or trusts information provided by an automated decision aid and the degree to which that system is in fact trustworthy.

References

Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1-73.

Kirby, K. N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition*, 51, 1-28.

Klayman, J. & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94(2), 211-228.

Latest News on Brunswikian Music Psychology

Patrik Juslin
Uppsala, Sweden

Music is probably the most widely practiced and appreciated of all art forms. One possible explanation for this may be that music offers a powerful means of emotional communication. Knowledge is scarce on this issue, however, in particular when it comes to performance of music. It may be argued that this partly stems from a lack of relevant theories. In my doctoral dissertation (Juslin, 1998a), I proposed a theoretical framework, the Functionalist Perspective, that integrates ideas from research on emotion and nonverbal communication with Brunswik's Lens Model. The usefulness of this framework was illustrated in three studies.

The first study showed that professional guitar players were able to play a piece of music so as to communicate specific emotions (i.e., happiness, sadness, anger, fear) to listeners. Acoustical analyses revealed that the performers used a number of probabilistic but partly redundant cues in the performance (e.g., tempo, sound level, and articulation) to generate the desired emotional expression (Juslin, 1997a).

The second study showed that synthesized performances based on the empirical data of Study I yielded predicted judgments of emotional expression from listeners. It was further shown that listeners used tempo, sound level, articulation, frequency spectrum, and tone attacks in their judgments. Linear regression models provided a good fit to cue utilization, and the cues contributed independently to judgments (Juslin, 1997b).

The third study used multiple regression analysis to describe cue utilization of both performers and listeners. The two systems were then related by means of the Lens Model Equation. The results showed that (a) about 80 percent of the variance in the listeners' judgments could be explained by the performer's expressive intention, (b) the accuracy of the communication depended mainly on the extent to which the cue weights of the performer "matched" the cue weights of the listener, (c) cue utilization was more consistent across pieces of music than across performers, and (d) there were cross-modal similarities in code usage between music performance and vocal expression of emotion. It was thus suggested that performers are able to communicate emotions to listeners by using the same acoustical code as is used in vocal expression (Juslin, 1998b).

I have stayed busy in the laboratory during the summer, working on two new projects with a Brunswikian flavor. The first project is an attempt to use Cognitive Feedback to improve the expressive skills of novice performers (Juslin & Laukka, in preparation). The background of this project is that it recently has been found that music teachers often fail to address expressive aspects of music performance, instead concentrating their time and effort on technical aspects. One reason for this may be that emotions are expressed and recognized in (mainly) implicit ways. Thus, teachers may find it difficult to verbalize many aspects of their expressive skills. This view is supported by studies suggesting that the feedback that teachers give their students is too vague to allow for the improvement of the students' expressive skills. In our study, we provide novice performers with a chance to compare their own cue utilization to the optimal model based on listeners' cue utilization. This involves a (distressingly) complicated pretest-posttest control group design which combines between-subjects and within-subjects measures to evaluate the efficacy of the feedback. We use both behavioral criteria and reaction criteria for the assessment, and we also measure the performers' policy insights prior to feedback. The study is still underway, but we are pretty confident that CFB will have positive effects on performers' expressive skills.

The second project concerns the importance of timing patterns in communication of emotion through piano performance (Juslin & Madison, in preparation). Through digital re-synthesis, we have been able to gradually eliminate various expressive cues from musical performances in order to see how these cue reductions affect listeners' judgments of the emotional expression. The preliminary results suggest that (a) listeners do use timing patterns to decode emotional expressions, (b) timing patterns are less effective in communicating emotions to listeners than are tempo and dynamics, (c) elimination of timing patterns does not necessarily reduce listeners' decoding accuracy (due to redundant information provided by other cues), (d) fear expressions and happiness expressions are more dependent on timing than are anger and sadness expressions, and (e) timing patterns alone are capable of communicating fear with better than chance accuracy. Idiographic analyses of the listeners' judgments of both impaired and unimpaired piano performances could yield further insights into the judgment policies of the listeners. This knowledge, in turn, could prove to be useful in music education.

Judgment research at Vanderbilt

Jim Hogge
Nashville, Tennessee

Steve Schilling and I are putting the finishing touches on "Modeling Diversity in Judgment: Hierarchical Linear Models as a Bridge Between Generalizability Theory and the Lens Model Equation" (promised in the Fall 1997 issue of The Brunswik Society Newsletter). Those interested in receiving a copy should alert me at the email

address given at the end of this note. Steve and I are also working on a chapter (for *The Essential Brunswick*) that will focus on the use of hierarchical linear models for the nomothetic aggregation of idiographic descriptions of judgment.

I am also preparing (with Jacqueline Palka) to collect data in a study of judgments of the professional competence of student teachers by university faculty, supervising teachers, and student teachers themselves. Specifically, we plan to examine (a) the relationship between self-insight and professional competence (as judged by supervising teachers) and (b) the relationship between other-insight and professional competence. We predict that students with high other-insight will receive higher ratings than students with low other-insight and that students with high self-insight will receive higher ratings than students with low self-insight. Also, we expect that the relationship between other-insight and professional competence ratings will be stronger than the relationship between self-insight and professional competence ratings.

Finally, I am preparing two additional chapters for *The Essential Brunswick*: (1) a description of the application of the lens model to the assessment of professional competence and (2) the use of generalizability theory to assess the reliability of expert judgment.

Performance of Physicians at Judging Survival in Congestive Heart Failure

Wally R. Smith
Richmond, Virginia

Roy Poses, Donna McClish, and I have been investigating how physicians make judgments of the probability of important outcomes, especially for the common clinical problem of congestive heart failure (CHF), a clinical syndrome characterized by the inability of the heart to pump blood adequately for the body's needs. Acute CHF may be a medical emergency and may result in a decision to admit a patient to an intensive care unit (ICU). Current guidelines suggest that physicians should base ICU triage decisions in part on their judgment of the probability that the patient will survive in the short-term (and imply that patients with a very small likelihood of survival should not be admitted to an ICU because care for them there would be futile.) We have shown that physicians' judgments of survival for patients with acute CHF made at the time the triage decision has to be made are poorly calibrated (overly pessimistic) and have minimal discriminating ability. So our next questions were how do physicians use relevant clinical cues when making these judgments, how well does a model of their judgments based on such cues predict survival, and how well do such cues actually predict survival.

We enrolled a sequential cohort of patients visiting Emergency Departments (ED's) at one of three hospitals, an urban university hospital, a VA hospital, or a community hospital in one metropolitan area. We excluded patients with acute myocardial infarction (or "heart attack," who are physiologically and clinically different from other patients with acute CHF), and excluded patients who died or developed an acute complication requiring ICU care in the ED (because there would have been no question about the triage decision for them.)

We collected data about clinical cues from a chart review, about survival from multiple sources, and about physicians' judgments prospectively at the time of the ED visit using a standardized instrument. We selected as cues variables that previous research or our clinical judgments suggested might be related to survival for patients with acute CHF.

We modelled the relationship of these variables to 90-day survival using logistic regression (survival model), and the relationship of the same set of variables to the logit of the physicians' survival judgments (judgment model) using linear regression.

The R squared for the judgment model was .20. The area under the ROC curve for the survival model was .76. Of the eight variables that independently predicted judgments or survival (Table), one predicted only judgment, five predicted only survival, and two predicted both.

Variable	p, Judgment	p, Survival
Age	.0001	.0007
Sodium	.9982	.0003
Low systolic blood pressure	.0602	.0411
Orthopnea or paroxysmal nocturnal dyspnea	.7608	.0002
Acute Coronary Artery Disease symptoms	.0120	.6402
Prior requirement of ACE inhibitor	.1771	.0281
Charlson comorbidity score	.2217	.0138
Functional status (ED judgment)	.0001	.0001

Physicians' judgments of survival for patients with acute CHF may be inaccurate because they fail to use cues that predict survival while using others that do not predict survival. Developing better predictive models and teaching physicians how to use them may improve clinical prediction and thus clinical decision making.

Research at University College London

Clare Harries
London, United Kingdom

1. Judgmental combination of Forecasts

Nigel Harvey and I are looking at the factors that affect our ability to combine other people's forecasts. We are doing this within the framework of a lens model, treating the forecasters as the cues, and modeling both the (artificial) environment and the judgmental combination of these forecasts. This work is part of a larger on-going project investigating judgmental combination of forecasts.

This year we began by investigating the effects of forecasts that exhibit typical patterns of behavior (such as a "trend bias"), on both explicit knowledge of the accuracy of those forecasts, and on tacit judgmental combinations. People seem to have a good explicit understanding of the validity of forecasts but are relatively insensitive to their intercorrelations. We found that forecasters were tacitly relied upon most ideally when they exhibited behavior that was atypically biased. We extended the research in two directions. In an investigation of the role of tacit and explicit forecasts we have looked at judgmental combination when one of the cues is in fact your own forecast, and which may or may not be labeled as such. In a further investigation of understanding of interforecaster correlations we have compared situations where forecasters tend to agree, and situations where they are equally accurate but do not tend to agree.

2. Models within the Lens Model Framework

At the SPUDM 1997 conference Ulrich Hoffrage demonstrated that a multiple linear regression did not distinguish between simulated judgements based on linear regression and those based on "take the best". In a paper delivered in December 1997 to the London J/DM group, Mandeep Dhimi and I discussed the relative merits of using a "Take the best" model rather than a regression model to describe the judge and the environment within a lens model framework. We compared the "Take the best" model with the classic findings using regression models in terms of fit, agreement, cue use, consistency and self-insight. Since then we have tested our theoretical conclusions through reanalysis of physicians' decisions. We hope to present the results at this year's Brunswik meeting. If this does not happen, e-mail us to find out more.

3. Occupational Therapists' decision making

Earlier this year Cilla Harries (an Occupational Therapist) and I presented a paper to the 12th International Congress of the World Federation of Occupational Therapists. Cilla has looked at OT reasoning using traditional qualitative methods, eliciting the OT's concurrent verbal reports carrying out in-depth interviews based on a handful of cases. Together we argued for increased use of quantitative methods in OT, as a compliment to the currently used qualitative methods. Quantitative methods in OT are likely to be associated with reductionist experimental methods. In our paper we introduced judgement analysis and described its advantages in relation to other types of modeling. We described what would have been gained by extending a study of OT referral decisions to include judgement analysis. We argued that by using interviews, concurrent verbalization and judgement analysis we could obtain a fuller description of the cues that people attend to, the ways the cues are interpreted, and an objective measure of the ways in which cues influence decisions.

Brunswik and Complexity

Ray W. Cooksey
Armidale, NSW, Australia

My research is evolving on several fronts. First, in conjunction with a postgraduate student, I am involved in a more traditional judgment analysis investigation of personnel manager judgments of job applicant suitability for interview. Identifying specific cues employed by managers making such judgments are a key focus of the research. Second, one of my PhD students has embarked on a unique investigation of the decisions and judgments made by private art collectors in Australia. Her interest is in retracing the decision processes involved in the pursuit and acquisition of specific art pieces by collectors. The research will involve a combination of think-aloud process tracing and judgment analysis. She has access to some of Australia's highest profile collectors and the analysis should provide some interesting insights into the behavior of art collectors.

I reported last year that I was engaging in some work that was not strictly Brunswikian, yet was founded on some very fundamental Brunswikian principles. This work continues. It focuses on establishing and testing complex dynamic systems conceptualizations of work behavior and decision making. Conceptualizations have been separately established for understanding work performance and for the decision making of court magistrates in light of highly interconnected environmental, organizational, interpersonal, and intrapersonal contexts. The work performance conceptualization has been tested in a small-scale idiographic study of teachers in a private Sydney high school. By tracking teacher perceptions of five interacting subsystems of influence on their work performance over time, the influences were shown to behave in a complex and chaotic manner that could not be captured by traditional linear modeling techniques. Work on testing the implications of the magistrates' courtroom decision-making conceptualization will likely proceed later this year, commencing with some in-depth interviews of a sample of NSW magistrates. I have also begun to merge some of this conceptual dynamic systems work into an MBA I teach on managerial decision making.

"The Essential Brunswik: Beginnings, Explications, Applications"

**Ken Hammond, Boulder, Colorado,
Tom Stewart, Albany, NY**

For some time we, as well as others, have been concerned that the core of Brunswik's work-the content of his original papers-will disappear in the mists of the past. Very few scholars have seen, let alone have read, more than a tiny fraction of Brunswik's papers. As a result, most of our knowledge and beliefs about his ideas have come from secondary sources. This is regrettable for several reasons; one is the loss of the true sense of the fundamentals of his general theoretical and methodological work; another is the absence of acquaintance with the manner in which Brunswik expressed his ideas, and that is truly to miss the maturity and breadth of his scholarship. So we think this situation ought to be remedied by publishing a volume of Brunswik's collected papers and applications of his work.

If we did only this, researchers and students could at least find his papers and read them, and thus gain knowledge of what he had to say in his own words. We think that would be valuable, but not enough; it would simply be an archival volume to be used as a resource. More is needed. What is needed is for scholars who know Brunswik's work to enlighten new researchers, and to show them meaning and significance, in the past and in the present, of these articles. In short, we want to make Brunswik's work come alive-as it deserves to-for psychology today. We have given a great deal of thought to the nature of this volume and what its objectives should be. Given the ever-increasing interest in Brunswikian ideas and the ever-expanding scope of the theory and research that has resulted from them, we have decided that a volume that merely included his papers with commentaries would not do justice to the results of his work. Consequently we have organized the volume so that it will reflect the impact of Brunswik and the work that followed, and still follows, from his writings. The volume will have three major parts: Part I -- Beginnings: The Grand Ideas Introduced; Part II -- Explications: Iconoclasm at Work; and Part III - Applications.

Oxford University Press has agreed to publish volume entitled "The Essential Brunswik: Beginnings, Explications, Applications" edited by Kenneth R. Hammond and Thomas R. Stewart. We are very pleased that OUP has recognized the importance of Brunswik's work in this way and that approximately 30 distinguished scholars have agreed to contribute to the volume. The publication date is late 1999.

Judgments under stress

**Kenneth Hammond
Boulder, Colorado**

This year has been devoted to working on my book ms "Judgments Under Stress", in which I undertake to present a Brunswikian approach to this, in my view, badly treated topic. Badly treated, that is, by clinical psychologists, and largely ignored by J/DM researchers. I produced a reviewable ms which has received good reviews and is at this writing being presented to the Executive Board at Oxford University Press. It is a fairly brief ms (about 250 typewritten pp) plus a rather long appendix which includes an annotated bibliography of research on judgments under stress. I had to include such a bibliography because my ms is iconoclastic; it argues that present research cannot be used because it lacks any organizing principle, therefore I had to show that I had indeed read the literature. Do I offer an organizing principle? Yes, of course. What is it? Well, I had to begin with a non-arbitrary definition of stress because it is precisely because all current definitions are arbitrary that progress is impossible. I defined stress in terms of a loss of constancy, both correspondence constancy and coherence constancy, and everything just followed from there. The idea of loss of constancy as

the origin of stress came to me because constancy is the principal focus of Brunswikian psychology; achievement of constancy might be said to be the glue that holds us all together. I am always surprised that psychologists do not make more out of this phenomenon; I believe it to be the fundamental discovery of psychology. (Am I wrong? Is there a contender?)

In any event, that is the key idea of the ms.

Some will be surprised to find that I give even more credence to documented reports of behavior under stress (loss of constancy) than I do to experimental evidence. There are lots of reasons for this but you will have to wait for the book appearance to learn what they are.

Policy Capturing and Multiple Regression: A Married Couple?

Ulrich Hoffrage
Berlin, Germany

Judgment analysis mainly uses linear models, particularly multiple regression, to describe how participants utilize available cues to arrive at their judgments. Although neo-Brunswikians have mostly restricted themselves to using multiple regression as a tool for describing judgments, this analysis is in principle open to testing other candidate models. Among these other models are fast and frugal heuristics, such as Take The Best (Gigerenzer & Goldstein, 1996, *Psychological Review*, pp. 650-669; Gigerenzer, Todd, and the ABC Research Group, in press, *Simple heuristics that make us smart*, Oxford University Press). Take The Best is designed for pair comparison tasks: If the most valid cue discriminates between two objects, the heuristic will choose the object the cue favors; if the most valid cue does not discriminate, the next best cue is checked, and so on.

I recently investigated the following question: How powerful is standard policy capturing (i.e., when based on multiple regression) in discriminating between several strategies that generated choices in pair comparison tasks? Different strategies frequently lead to the identical prediction so that an individual's decision cannot be assigned unambiguously to one strategy. To illustrate this problem, the overlap between the predictions of Take The Best and those of multiple regression has been determined for the environment of German cities used by Gigerenzer and Goldstein (1996). It turns out that in 96% of the comparisons, where both Take The Best and multiple regression made a prediction, their predictions were identical. Such a large overlap has implications for policy capturing. Suppose that a participant consistently uses Take The Best to make inferences for 100 paired comparisons randomly drawn from this city environment and that another participant consistently uses multiple regression for the same set. Would policy capturing detect any difference between these two participants? The answer is no. The multiple correlations and the weights in the regression equation that describe the policies of these two participants do not differ (much) from each other. One solution to this problem of separability, that is, the difficulty of discriminating between strategies, is to select the alternatives presented to participants in a way that forces the strategies to make different predictions. However, selecting alternatives to minimize the overlap of the strategies' predictions often makes the item set unrepresentative and the results difficult to generalize. There seems to be a dilemma here: Either the item set is representative, with generalizable results but barely distinguishable strategies, or the item set is selective, with distinguishable strategies but possibly limited generalizability.

The idea of a fast and frugal lens model seems to be worthy closer inspection and this is a direction I will take in further research.

Thus, some of my other research interests and activities that are listed below also deal with these simple heuristics (to avoid a wrong impression: in most cases I'm the co-author):

- Ecological rationality of fast and frugal heuristics: How does their performance depend on the structure of the information in the environment? (with Laura Martignon)
- Effect of time pressure on judgment and decision making (with Jörg Rieskamp)
- Intransitivities and rationality (with Martin Lages)
- Fast and frugal heuristics for quantitative estimation (with Ralph Hertwig)
- Hindsight bias as a by-product of an adaptive process (with Ralph Hertwig)
- Bayesian Inferences and representation of information, including studies with experts, e.g., with physicians, or AIDS counselors (with Gerd Gigerenzer)
- Risk taking behavior of children in traffic (with Angelika Weber)

Overconfidence, causal systems, and other people's feelings

Joshua Klayman
Chicago, Illinois

I am working on three loosely Brunswikian projects.

1. A revival of a very old project on overconfidence with Jack Soll and Claudia Gonzalez-Vallejo. We finally figured out (we think) what to make of some data we collected several years ago. Using new analytical techniques, we find that there is indeed some overall bias toward overconfidence, but it varies greatly with what domain you are asking about, how you ask (confidence in a two-choice question or setting a confidence interval), and who you ask (i.e., there are stable individual differences). We do not find any evidence for an effect of difficulty on overconfidence when we control for the kinds of confounds Gigerenzer and colleagues warn about.

2. Continuation of a middle-aged project with Alex Wearing. We have run two studies now on how people learn the causal relations in interconnected causal systems. Using simple three- and four-variable systems, our preliminary findings are that people are pretty good at picking up the correlational structure of the system, but they have trouble distinguishing the possible causal relations underlying the covariations they observe. In particular, people tend to see direct causation where there is only indirect, and causal links where there are only "spurious correlations." With grad student Boris Brodsky, we also have preliminary data on people's ability to think of causal loops as explanations. I think we'll go on to examine what strategies of experimentation and observation are most or least effective for extracting the causal structure from observed covariations in systems.

3. Early work on a new project with various people perhaps including Chip Heath, Chris Hsee, and George Loewenstein. We're thinking about the cues people use in trying to figure out other people's feelings, reactions, preferences, etc. (We think very similar processes may also apply to predicting one's future self and explaining one's past self.) We will try to tie together a variety of findings in cognitive and social psychology by

suggesting that there is a consistent pattern of errors that follows from the cues people have available to them, and how they fill in missing or unreliable information.

Simulation, stress, and visualization research in Georgia

Rob Mahan

Athens, Georgia

For the last several months we have been busy building an internet-based AWACS simulator for the Air Force. This has been a very labor intensive process because we decided to build it from the ground up. Our goal is to create a platform independent system, yet be able to offer a degree of fidelity that is found in large-scale device-based training environments. Selecting JAVA as our programming language has proven to be both good and not so good. Many of the programming features we need in order to implement a variety of high fidelity components of the sim have yet to be developed. Our AWACS interface includes A/V channels, Virtual Reality 3d visualization and other components not found in simulations of this type. We have been using special beta code from Sun Microsystems to test some of these concepts and are making progress. Nick named SynTEAM by the Air Force (synthetic team effectiveness assessment and modeling), the sim focuses on examining and eventually training AWACS teams in a cost effective manner (at least in theory). Of course, our vision of a performance measurement system includes a heavy dose of Brunswikian philosophy and concepts, and we have made more than a few Air Force types anxious for insisting that we have a way to this [e.g., embedding hierarchical lens model features into the simulation]. We are also kicking off a grant next month to begin building and integrating an intelligent coaching system into the simulation as well.

On our much-neglected empirical side, we continue to persevere examining the effects of stress (e.g., fatigue, sleep deprivation) on judgment performance. We are writing up a paper on the effects of sleep deprivation and fatigue on team hierarchical judgment. We find that during 24 hours of sleep loss, both cognitive control AND matching deteriorate. However, the strategic changes that occur in the teams are particularly interesting, and these are related to voluntary shifts in cognitive mode. This paper will be submitted to the journal Aviation Space and Environmental Medicine in the next month or so for a stress issue we are editing.

Related to our stress and judgment work, we are examining the efficacy of visualization (representation) techniques that can drive an operator to a particular cognitive mode. We have finished two papers now that show, in part, that the characteristics of the display can force an individual to process in a particular ways (Intern. Journal of Cognitive Ergonomics in October; the other submitted). While display induced changes in processing is nothing new, using the Cognitive Continuum Theory as a way to understand these changes is (to the best of our knowledge). We have been using the Cognitive Continuum as the framework to predict and explain these shifts in cognition, and how one can leverage display dynamics to achieve specific (efficient) modes of cognition. Our interest in this area emerged out of some of our stress work that showed people begin to drift on the continuum when fatigued. Our goal here, of course, is to create ways (information packaging protocols) that might serve as countermeasures against the effects of fatigue and sleep deprivation.

Learning and performance in interactive dynamic decision tasks

Bud Gibson

Ann Arbor, Michigan

My research focuses on understanding factors that affect learning and performance in interactive dynamic decision tasks such as telephone based credit collections or stock trading. The work is multidisciplinary. I use computational modeling and lab experiments to build and test theories about learning. Field studies provide an opportunity to ground these two activities. This Fall I plan to begin developing experiments around a simulated credit collections task which I want to use to study the effect of fear appeals on interactive decision making.

Two recent papers provide a more in-depth review of the work, recent results, and implications for decision making which I believe may be of interest to members of the Brunswik Society. I have included the abstracts as part of this research summary. Interested parties should feel free to contact me directly (fpgibson@umich.edu) if they would like copies once the papers are ready for wider release. The first paper is:

Learning in Dynamic Tasks with Feedback Delays

Decision makers in dynamic environments (e.g., stock trading, inventory control, and firefighting) learn poorly in experiments where feedback about the outcomes of their actions is delayed. In an effort to generate ways to mitigate these effects, this paper presents two computational models of learning with feedback delays and contrasts them against human decision makers' performance. The no-memory model hypothesizes that decision makers always perceive feedback as immediate. The with-memory model hypothesizes that, over time, decision makers are able to develop internal representations of the task that help them to perform with delayed feedback. As borne out by human subjects, both models predict that a representation of past history improves learning with delay and that increasing delay increasingly degrades performance. Even though the length of training in this task exceeds that used in many laboratory-based dynamic tasks, neither the two models nor the subjects are able to effectively learn without decision aids when faced with feedback delays. When given an amount of training that may more closely approximate that provided in functioning dynamic environments such as telephone-based credit collections, the with-memory model predicts that human decision makers may learn without decision aids over the long-term when feedback delays are simple. These results raise several issues for continued theoretical investigation as well as potential suggestions for training and supporting decision makers in dynamic environments with feedback delays.

The second paper is:

Fear appeals: What happens when the credit collector calls?

When fear appeals are used to gain compliance, the target is asked to do something for which he or she has disincentive in order to avoid a worse consequence. Much emphasis has been placed on how the emotional and information content of messages concerning these consequences might be manipulated in order to raise the rate of compliance. We examined 192 contacts between credit collectors and debtors to understand the effectiveness of fear appeals in a functioning organization. Surprisingly, messages concerning the consequences of non-compliance have inconsistent or insignificant effects. Our results suggest that collectors are most effective with delinquent debtors when they focus on the specifics of what is required for compliance. This task focus during the interaction shows debtors a clear way to either minimize their negative emotion or, alternatively, to grapple with the dangers the fear appeal alerts them to. This result may have broader application in sales and other interactive boundary-spanning work roles where emotional displays have been considered important to effective performance.

Accuracy of physician judgment

Neal Dawson
Cleveland, Ohio

Our research this past year has used methods that we believe are valuable in performing nomothetic comparisons in real life medical care settings.

We have been evaluating the accuracy of physician judgments in actual clinical settings and have compared the accuracy of easily identified groups of physicians, e.g., physicians from different medical specialties. This has led to the need to address the issue of physicians seeing similar but not identical patients in clinical settings, e.g., patients who have the same diagnosis but may have differing levels of severity of illness or comorbidity. This is caused by selection bias. In prior studies of factors that influence patient survival, we have used the propensity score methodology to adjust for selection bias that arises when patients are selected (e.g., for a given treatment) or are otherwise nonrandomly distributed within a clinical care setting based on patient characteristics (e.g., severity of illness). A propensity score for a dichotomous outcome is created using logistic regression. The dependent variable is the outcome of interest (e.g., survival) and the independent variables are the patient based characteristics that are associated with selection (e.g., for treatment) or that are associated with the outcome of interest (e.g., mortality). We have adapted this methodology to create groups of patients who have sufficiently similar baseline characteristics to allow a fair comparison of judgmental accuracy.

We were interested in comparing the accuracy of survival estimates of two types of physicians who care for seriously ill cancer patients: generalists and oncologists. Using patient characteristics, we created a propensity model that predicted the likelihood of a patient seeing an oncologist (vs. a generalist). The adjustment for selection bias created two groups of essentially identical patients with respect to baseline characteristics that are important to mortality risk. We then compared the accuracy of generalists and oncologists judgments of survival of patients under their care.

Across all judgments, we found that oncologists were generally too optimistic in their predictions and that generalists were too pessimistic but were more accurate than the oncologists. Accuracy varied by patient age such that oncologists were more accurate for younger patients (under 40), generalists were more accurate for middle aged and for older patients. Accuracy also varied by the number of estimates made (=10). This relationship is currently being evaluated and will be presented at the Brunswik Society meeting in Dallas along with details of the propensity score methodology.

Current activities of a more or less Brunswikian nature

Alexander J. Wearing
Melbourne, Australia

1. With Mary Omodei and Jim McLennan we have been continuing the development of, and experiments with Networked Firechief, This program allows multiple persons, each with their own computer, to collaborate in fighting a simulated fire (or carry out an analogous task such as dealing with an oil spill or a locust plague). We are interested in the consequences of constraints on command and control, and the role of information, time pressure, command authority, and person characteristics in performance.

2. Related research has involved field work with firefighters who are wearing helmets into which have been built small video cameras and microphones for recording the own-point-of-view experience of the firefighters, and using the resulting information in naturalistic decision making studies.
3. With Leon Mann, we have been investigating the extent to which decision making style, measured as a trait, relates to organizational variables such as organizational climate and leadership, as well as other person characteristics.
4. With Josh Klayman , we have a continuing (in a leisurely way) project. We have run studies now on how people learn the causal relations in interconnected causal systems.
5. Oswald Huber and I are bringing to a close a project that involved developing and testing models that attempt to simulate the behavior of subjects on a multi-stage investment task.
6. A minor line of research has been on economic judgments; in particular, what variables or cues influence the evaluation of different taxation packages.