



The Brunswik Society Newsletter

www.brunswik.org

Edited by Clare Harries

October 2002

Tom Stewart is finally recognized – p.2

Obituaries: Mats Björkman and Derick Steinman – p.3

Eighteenth Annual Meeting of the Brunswik Society - p.27-28

Research news (Page, Author, Title)

4	Adelman	Team Decision Making Under Time Pressure
4	Athanasou	The Role of Contextual Factors in Repeated Judgements: Implications for Workplace Learning
5	Anderson	The Root of Error: Unreliability as a Response to Task Environments
5	Bass	Human-automated Judgment Learning: A Methodology to Investigate Human Interaction with Automated Judges
6	Bisantz	Research on Visualisation and Decision Making
7	Chasseigne	Aging and Probabilistic Learning: Further Evidence
7	Denig and Mol	Decisions to Prescribe Antimicrobial Treatment
8	Dhami	Effects of Systematic and Representative Stimulus Design on Policy-Capturing
8	Dibadj	Statistical Properties of the Parameters of the Lens Model Equation
9	Doherty	Study Design, Diagnosis and Probability Judgments and Brunswik's Contribution to the History of Research on Thinking
9	Dowding, Paterson & Cassells	Risk of Suicide and Decisions to Observe Psychiatric In-Patients
9	Dunwoody	The Use of Base Rate Information as a Function of Consistency and Diagnosticity
10	Funder	Towards a Balanced Social Psychology
10	Galbraith	Assessing Clinical Decisions of Medical Students
10	Gattie & Bisantz	Applying Real-time Cognitive Feedback in a Dental Diagnosis Task
11	Gifford	Brunswikian Architecture
11	Gonzalez Vallejo	Are All Judgments Equal?
11	Gunderson	Using Judgment Analysis to Predict the Targets of Crime.
12	Hamm	Prostate Cancer Screening: The Lens Model Clarifies a Comparison between the Health Beliefs Model and a Descriptive Expected Utility Model.
12	Hammond	Progress?
13	Harries	Information Search, Use, Reasoning and Being strategic (but never all at once)
13	Harvey	Advice Taking and Trust in Advisors
14	Holzworth	Brunswikian Research at the University of Connecticut
14	Jha, Bisantz & Parasuraman	Applying Lens to Free Flight: A Lens Model Analysis of Pilot and Controller Decision Making in a Future ATM System
15	Juslin	Exemplar Matching and Cue Criterion Relations
15	Kerstholt & Koele	Tracing Shoe Tracks
16	Kirlik	Toward a Coherent Ecological Perspective
16	Koele	Configural judgement Strategies and MCPL
17	Klayman	Different Kinds of Confidence
17	Kurz-Milcke	More on How he Developed his Ideas
18	Leavitt	Representative Design and Social Perception
18	MacCormick	Clinical Judgment Analysis of Patient Prioritisation for Elective General Surgery
19	Mosier	Coherence and Correspondence in an Automated Flight Deck
20	Nath	A Validation Analysis of Methods of Utility Estimation
20	Newell and Shanks	Predicting Search Strategies in Simple Heuristics and Using Feedback in MCPL
21	Pijnenburg	Monitoring and Aiding Clinical Decision Making under the new Dutch Youth Care Act
21	Ramos Alvarez	The Importance of the Task Environment in Processing Threatening Stimuli
22	Ravitch	Decisions to Prescribe Estrogen Replacement Therapy
22	Sarver, Dawson et al	Rapid Clinical Decision in Context: A Theoretical Framework to Understand Physician Decision-Making
22	Seong & Bisantz	Operators' Judgment Performance with Automated Decision Aids: A Theoretical Model and Empirical Investigation
23	Shanteau et al	Can Decision Errors Be Predicted Before They Happen? Application of a New Measure of Skilled Performance
24	Skå nér	Diagnosing Chronic Heart Failure in Primary Health Care
24	Smith	General Practitioners' decision making and the role of clinical guidelines
25	Sorum	Treatment vs Diagnosis
25	Tweney	Further Studies of Epistemic Artefacts
25	Weaver	Persuasion, Judgment and Learning
26	Wigton	Physician Diagnosis and Management of Respiratory Tract Infection
26	Wittmann	Brunswik-Symmetry

THE BRUNSWIK SOCIETY

Recognizes

Thomas R. Stewart

**In appreciation for many years of dedicated service to The
Brunswik Society**

**For outstanding scholarly contributions in expert judgment and
forecasting that have significantly advanced Egon Brunswik's
vision**

**For steady dependability, gracious manner of stewardship,
unfailing friendship, and collegiality**

**For low key, wry humor and encouragement to everyone around
him**

November 16, 2001

Obituaries

Although several months have passed, this is the first newsletter since the deaths of Derick Steinmann and Mats Björkman. I have reprinted the simple notices that were sent to announce the sad pieces of news. More information is on the Brunswik website at www.brunswik.org.

DERICK STEINMANN

Dec. 5, 2001

Mike Doherty wrote:

It saddens me to inform the members of the Brunswik Society that Derick Steinmann passed away today at his home in the Virgin Islands. Derick did his dissertation under my supervision at Bowling Green in 1972, after which he spent two years on a post doctoral fellowship at Boulder with Ken Hammond. His first foray into judgment research was a 1972 paper in OBHP, "A Lens Model Analysis of a Bookbag and Poker Chip Study." Derick's dissertation, "Transfer of Lens Model Training," was published in OBHP in 1972, but the paper that is most closely associated with his name is the 1975 paper in the Kaplan & Schwartz volume with Ken Hammond, Tom Stewart and Berndt Brehmer. After Derick left Boulder he went into the business world, ultimately buying and selling a very successful newspaper. Coincidentally, one of his two sons lives in Boulder. My memories of Derick are good ones; he was a good person.

Mike Doherty

MATS BJÖRKMAN

19/11/2001

Peter Juslin wrote:

Dear Colleagues,

Last Friday, 16th of November, my friend, colleague, and former PhD-thesis supervisor Mats Björkman (1926 - 2001) passed away after a long time of serious illness. As many of you will know, Mats Björkman was an early proponent of Brunswikian ideas in Psychology, establishing them as one of the core themes of Swedish Judgment and Decision Making research in the sixties. Aside from his original contributions to JDM research, he leaves a long string of students and "grand-students" behind him that have promoted these ideas further. Of course, to us who had the fortune to know him personally the loss is not primarily professional but personal: Mats was a great friend and human being.

Peter Juslin

Team Decision Making Under Time Pressure

Len Adelman
 George Mason University
 E-mail: ladelman@gmu.edu

My students and I continued to perform research on the effect of time pressure on decision making for a simulated team, air defense task. Without going into the details, we found the reward structure to be much more effective than interface features in maintaining high levels of decision accuracy under high levels of time pressure. The cost for maintaining accuracy was that operators made fewer decisions and sent less information. These results, combined with our previous research, suggest that, depending on the team decision making task and support environment, (1) there is a time pressure level beyond which operators can not maintain both decision accuracy and quantity, and (2) if one wants to maintain accuracy, the reward structure and not the interface, may be the more effective mechanism for doing so. (By the way, a more technically detailed and expanded version of our Brunswik book chapter should appear in *Acta Psychologica* later this year. In addition, we published a paper in the March/April issue of *IEEE Internet Computing* describing an experiment showing that an actual website with alternative comparison features led people to use significantly more compensatory than non-compensatory decision strategies, while a site without such features did the opposite.)

The Role of Contextual Factors in Repeated Judgements: Implications for Workplace Learning

James Athanasou
 Faculty of Education
 University of Technology, Sydney
Jim.Athanasou@uts.edu.au

The purpose of this research is to test a new theory of learning proposed by Halliday and Hager (2002), who set out a relationship between context, judgement and learning (see Figure 1 for my representation of their ideas).

They saw learning as concerned with judgements that are potentially fallible but also contextually sensitive. Their approach was based on philosophical foundations and was not intended to provide a testable model. The objective of this research is to focus on the context aspects of the Hager-Halliday theory and the initial key research questions for this study are:

- To which aspects of a complex situation do people respond when they are learning?
- Can we model their construction of the situation?

The broader aim of this study is to lay the empirical foundations for a wider program of research that will focus on the antecedents, processes and consequences associated with workplace learning.

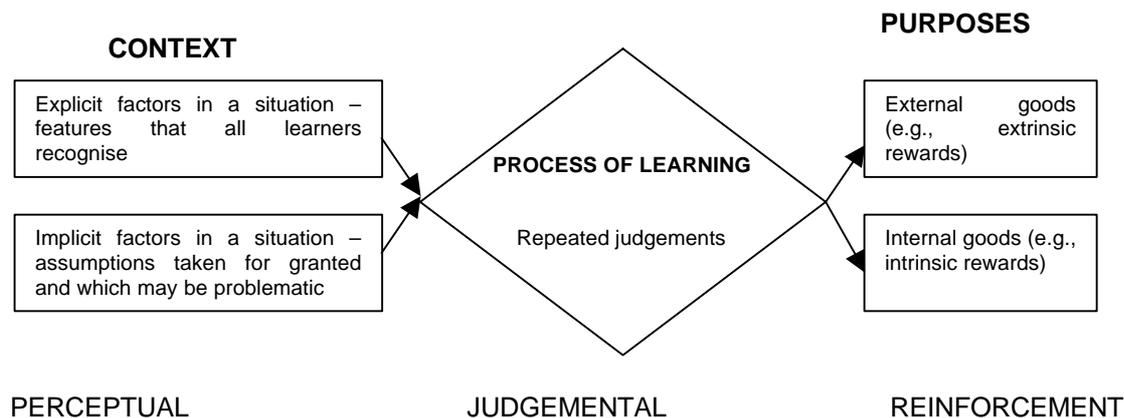


Figure 1. The Hager-Halliday approach expressed as a Perceptual-Judgemental-Reinforcement model of learning (note that the model is recursive)

Athanasou cont.

Background

Hager and Halliday (2002) argued that judgement is central to learning and that interests, purposes as well as features of a situation affect the judgement processes. The starting point for their model is the relationship between factors in a situation and judgements. As expected, the judgements are based on personally relevant features of a situation and in this aspect the model has major similarities to the field of research instigated by Brunswik on the importance of perception and judgement for all human responding. Hager and Halliday (2002) also hypothesised that judgements remain contextually sensitive to implicit and explicit features of a situation. They laid out a general plan but did not specify how these features interacted.

Research plan

Individuals will be exposed to situations in which firstly there is a single, identifiably correct solution and separately to an everyday situation where there is a subjectively preferred or personally optimal answer. If the theory proposed by Halliday and Hager (2002) is correct then people will repond to both implicit and explicit features of the situation in lawful but idiosyncratic ways.

The independent variables in this research are the cues (explicit and implicit) and the dependent variables are the lens model parameters (such as achievement, knowledge, unmodeled knowledge).

In this study three judgement situations that involve learning along a continuum of objectivity of outcome will be examined. The first relates to job choice (where there is no clearly identified correct answer but there are personal and subjectively optimal answers); the second situation relates to the correct medical diagnosis of systemic lupus erythematosus (SLE - the clinical features of SLE are protean and may mimic infectious mononucleosis, lymphoma, or other systemic disease). and the third situation is based upon spatial judgements involving shape analysis (where there is one correct answer). This project will be undertaken jointly by James Athanasou and Paul Hager.

References

- Hager, P. & Halliday, J. (2002). The importance of context and judgement in learning. In B Haynes (Ed.), Proceedings of the 30th Conference of the Philosophy of Education Society of Australasia. Churchlands: Edith Cowan University, pp. 15-30.
- Athanasou, J.A. (in press). The role of contextual factors in judgements: Implications for research into adult learning. Australian Vocational Education Review.

The Root of Error: Unreliability as a Response to Task Environments

Christopher J. Anderson and Thomas R. Stewart
E-mail: ca4809@albany.edu

In the past year, we have begun a new program of research that aims to expand our understanding of the causes of judgment error, in particular, the crucial and often overlooked role of human unreliability. We developed a theoretical framework based in the lens model that shows how reliability and consistency are related and how they contribute to reductions in achievement and potential increases in conditional bias. We have also conducted a meta-analysis demonstrating that task predictability is one source of inconsistency. This added inconsistency mediates a decline in relative achievement in low predictability environments. Thus, the environment produces error not just by setting a ceiling on accuracy, but also via its psychological effect on the judge. Finally, we conducted a series of six experiments that aimed to determine the variables that influence reliability, and in turn observed the effect these changes had on other judgment parameters such as relative achievement, bias, matching, and consistency. Generally, we have found that complexity led judges to be less reliable, which contributed to a notable decrease in accuracy in more complex environments. On the other hand, some very simple and inexpensive interventions were able to boost reliability significantly and thus increase accuracy. The work thus far suggests that investigating reliability provides an important window on judgment error. Our work also provides a foundation for more developed theories regarding the mechanisms that produce judgment reliability.

Human-Automated Judgment Learning: A Methodology to Investigate Human Interaction with Automated Judges

Ellen J. Bass
ejb4n@virginia.edu

Researchers in interpersonal conflict and interpersonal learning (IPL) have applied judgment analysis concepts and techniques in order to examine conflict between different judges working on the same task (Hammond, Wilkins, & Todd, 1966; Earle, 1973; Hammond, 1973). IPL consists of both an experimental process and a set of measures that investigate how one learns from the environment and another person as well as how one learns about another's judgment policy. Instead of studying how a person interacts with another human judge, Human-automated

Judgment Learning (HAJL) examines a person's interaction with and learning about an automated judge. HAJL is a methodology for investigating human-automated judgment system interaction capturing the judgment processes of the human and automated judge, features of the task environment, and relationships between them. As with IPL, it includes three phases: training, interactive learning, and prediction. In training, the human is trained to make judgments about the environmental criterion without interaction with the automated judge. In the interactive learning phase, the human judge first provides a judgment before having access to the automated judge's judgment and then provides a revised judgment after access. In the prediction phase, the human judge provides judgments with respect to the environmental criterion and predicts what the automated judge would judge. HAJL provides measures for conflict between the judges, compromise by the human judge, adaptation of the human judge to the automated one, and for assessing how well the human judge understands the automated one. HAJL was empirically tested using a simplified air traffic conflict prediction task. Two between-subjects manipulations were crossed to investigate HAJL's sensitivity to training and design interventions. Statistically significant differences were found with respect to 1) males outperforming females judgment performance before feedback from the automated judge was available while the automated judge's subsequent output eliminated this difference; 2) participants tended to compromise with the automated judge over time. HAJL also identified a trend for participants with higher judgment achievement to predict better the automated judgment and thought that their own judgments were closer to the automated judge than they were.

Research on Visualization and Decision Making

Ann Bisantz
University at Buffalo,
The State University of New York
bisantz@eng.buffalo.edu

Over the past year we have been involved in a number of studies involving human judgment and information visualization. Separate reports submitted by Pratik Jha, Gordon Gattie, and Younho Seong will present details of the our collaborative projects in areas of modeling pilot and controller judgments in air traffic management, applying real time cognitive feedback in a dental diagnosis task, and the use of Lens Model based framework to understand trust and calibration of use of automation systems.

In other studies, we are continuing work focused on methods for displaying probabilistic

information to decision makers, including research in this focused on developing and investigating the properties of graphical representations of uncertainty based on blurred or degraded icons, as well as visual, auditory, and tactile representations of spatially distributed uncertainty. The research has revealed through a number of studies that people are able to successfully map iconic representations to underlying concepts of uncertainty, and that performance on dynamic decision making tasks using such representations is similar to that of numeric representations. Further studies explored participants' interpretation of the icons by empirically generating fuzzy membership functions which mapped their interpretation of the icons' meaning to probabilities. Results from this work support the experimental findings, indicating that people generated membership functions with maximal values closely correlated to the intended numeric values. Additionally, it indicated that membership functions were reasonably similar across individuals. In combination, these are important findings, because it suggests that such representations may be implementable, and allow display designers to use a single icon to encode the uncertainty about an object's identity (e.g., whether it is hostile or friendly aircraft) along with other dimensions such as its location. These studies were supported by grants from the US Airforce Human Effectiveness Directorate, and the National Science Foundation (#IIS9984079), and were performed in collaboration with students Stephanie Schinzing, Jessica Munch, and Richard Finger. Additionally, a set of icons based on military symbology was generated for implementation in a demonstration battlespace visualization system, as part of a research effort funded by the Sarnoff Corporation. Also, in cooperation with co-investigator T. Kesevadas and M. S> student Santosh Basapur, research was performed to compare the utility of visual, auditory, and tactile modes for communicating a two dimensional probability density function. Specifically, color, tone pitch, and vibration were used to encode levels of uncertainty associated with points in a grid, representing probability of a hazard (i.e., an explosive or mine). Initial findings from a path finding task indicated that the visual modality resulted in path lengths that were less risky, but took longer than other modalities. We plan to pursue further work in this area to explore different schemes within each modality as well individuals' mappings of representations to levels of uncertainty.

Finally, we are currently undertaking studies to understand how people adapt their judgment strategies in cue-criterion judgment tasks, as the underlying probabilistic structure describing the cue-criterion relationships change,

and how different forms of cognitive feedback may affect that adaptation.

Aging and Probabilistic Learning: Further Evidence.

G. Chasseigne
François Rabelais University, France
E-mail: chasseigne@univ-tours.fr

We have been pursuing the research program initiated in 1997, aimed at examining the effect of age on probabilistic learning. In a MCPL experiment, Chasseigne, Mullet and Stewart (1997) examined the ability of elderly people to learn inverse relationships between cues and criterion. Three age groups (20-30, 65-75, 76-90 years old) of participants were asked to learn to predict the value of a criterion on the basis of three cues. In a first condition all cue-criterion relationships were direct (positive linear). Elderly participants were able to learn nearly as well as young participants, whatever the performance measure considered. In a second condition two cue-criterion relationships were direct but the third was inverse (negative linear). In this condition, learning was significantly worse in the elderly than in the young participants. The lower performance of the elderly people as compared with the younger people was essentially due to their inability to learn the inverse relationship. The only thing they learned was not to use the cue with an inverse relationship with the criterion.

In a third condition, subjects were given task information. The very elderly were not able to apply the knowledge of the inverse relationship provided by the cognitive feedback whereas the less elderly did not find it as difficult to modify their cognitive functioning. Chasseigne, Mullet and Stewart interpreted their result as a gradual decrease of cognitive flexibility in older adults.

In a recent study (Chasseigne et al., 2003), we examined further some of the conditions under which elderly people are able to learn probabilistic inverse relationships and when this type of learning is no longer possible. Two kind of tasks were used: (a) two single-cue learning tasks with either direct or inverse relationships (SCPL paradigm), and (b) three two-cue learning tasks, one with two direct relationships, one with a combination of direct and inverse relationships, and one with two inverse relationships (MCPL paradigm). Four groups of participants were included in the study: young adults (18-25 year-olds), adults (40-50 year-olds), elderly people (65-74 year-olds), and very elderly people (75-90 year-olds). It was shown that (a) older adults are able to reject the direct relationship "default" hypothesis and select the inverse relationship hypothesis when outcome feedback contradicting the default hypothesis is given, and provided that the

learning setting be a very simple one, involving only one cue; (b) some older adults are able to select the inverse relationship hypothesis provided that the learning setting be a simple one, involving only two inverse relationship cues; and (c) very few older adults are able to select the inverse relationship hypothesis when the learning setting is a complex one, involving two cues with both direct and inverse relationships with the criterion.

These results led to revise the "gradual decrease of cognitive flexibility in older adults" hypothesis proposed by Chasseigne, Mullet and Stewart. Moreover, there is a clear relationship between the Brunswikian probabilistic learning paradigms and the theoretical framework offered by the "executive function" construct. In the present study, the participants' task was to learn inverse relationships. They had (a) to inhibit the prepotent DR response and to substitute an IR response, and (b) to plan a sequence of processes for correctly estimating a criterion value from two predictor values. Inhibiting was an easy task for all participants when the situation was reduced to a single-cue one. Planning a sequence of processes for estimating a criterion was also an easy task for all participants when the situation only involved DRs. However, (a) when the situation involved a higher level of executive functioning, that is, when the participants had to inhibit a prepotent response and coordinating two cue values of equal meaning (IR and IR), most elderly people failed, and (b) when the situation involved a still higher level of executive functioning, that is, when the participants had to inhibit a prepotent response and coordinating two cue values of opposite meaning (DR and IR), all elderly people failed.

- Chasseigne, G., Ligneau, C., Grau, S., Le Gall, A., Roque, M., & Mullet, E. (2003). Aging and Probabilistic Learning in single- and multiple-cue tasks. *Experimental Aging Research*, 29 (in press).
- Chasseigne, G., Mullet, E., & Stewart, T. R. (1997). Aging and multiple cue probability learning : the case of inverse relationships. *Acta Psychologica*, 97, 235-252.

Decisions to Prescribe Antimicrobial Treatment

Petra Denig and Peter Mol
Department of Clinical Pharmacology,
University of Groningen, The Netherlands
E-mail: p.denig@med.rug.nl

We are conducting a number of studies related to decisions to prescribe antimicrobial treatment, both in general practice and in hospital. Evaluations of actualprescribing showed that adherence to antimicrobial treatment guidelines is low for urinary

tract infections and for sepsis. A combination of methods is used to study reasons for non-adherence. In general practice, think-aloud methods using vignettes showed that many treatment decisions are made without any contemplation of potentially relevant cues, suggesting that doctors often rely on simple decision routines. In some cases, this involves routinely prescribing second-choice drugs. This was recently published in *Qual Saf Health Care* 2002; 11: 137-43. In hospital, preliminary results from a lens model study using actual cases did not provide a good explanation for non-adherence. Factors included in the analysis were a range of potentially relevant and irrelevant case characteristics, such as age, gender, liver and kidney function, fever, leucocytes-counts, comorbidity, hospital acquired infection. It seems that many of the empirical antimicrobial treatment decisions are routinely made choices for second-choice drugs.

Effects of Systematic and Representative Stimulus Design on Policy-Capturing

Mandeep K. Dhani
University of Victoria, B.C., Canada
E-mail: mkdhami@uvic.ca

Judgment policy-capturing is perhaps one of the main avenues for neo-Brunswikian research. This literature suggests that among other things, judgments are the result of a linear, additive process, where few differentially weighted cues are used. The findings tend to be robust: applying to many different types of judges and judgment tasks. However, the fact that most studies involve hypothetical stimuli that comprise a factorial (orthogonal) combination of cues, draws attention to the questionable validity of the findings.

According to Brunswik (1955, 1956), stimulus design matters. The judgment policies that researchers try to capture are not adapted to the systematically designed stimuli that are presented to participants in the laboratory. Rather, policies are adapted to the stimuli that people naturally sample during their experience with the task. Brunswik believed this called for a new methodology: representative design. Representatively sampled stimuli may differ from systematically designed stimuli in terms of the number of cues, their values, inter-correlations, distributions, and ecological validities. Therefore, if Brunswik is correct, we should expect to find differences in the nature of the policies captured under systematic and representative conditions. Furthermore, we should expect that policies captured under systematic conditions would be poor at predicting judgments made on representative stimuli.

I conducted an experiment to test these two hypotheses. Senior college students were asked to judge the desirability of two sets of apartments for rent. One set included real apartments that were randomly sampled from an Apartment Directory given out to all students by the University housing office. These real apartments were described in terms of seven cues. The other set consisted of hypothetical apartments comprising an orthogonal combination of the cues. Overall, the policies captured for each individual differed: people 'used' fewer cues in the systematic condition, and the weights attached to the 'most important' cues differed under the two stimulus conditions. Furthermore, the policies captured using systematically designed stimuli were poor predictors of individuals' judgments on the representative set of cases.

These findings contribute to a small, but growing body of research demonstrating the effects of stimulus design on policy-capturing. Perhaps we need to sit back and think about what policy-capturing research over the past 50 years has 'really' told us about human judgment. The need for policy-capturing researchers to think about methodological issues is of great importance because of the impact their findings often have for professional judgment and social policies.

Statistical properties of the parameters of the lens model equation

Hassan Dibadj,
SUNY at Albany
E-mail: hd7554@csc.albany.edu

I am beginning work on a dissertation about exploring the statistical behavior of the parameters of the lens model equation and understanding how this behavior reflects the characteristics of the environment, the information system and the cognitive system. The goal of this study is to discover the statistical properties (distributions) of the parameters of the extended lens model equation under various assumptions about the judgment system and the environmental system being judged. The simulations will include factors and situations that affect judgmental performance as well as factors that affect the performance of judgment analysis techniques. The results of this study will facilitate the use and interpretation of the parameters of lens model equation.

Study design, diagnosis and probability judgments and Brunswik's contribution to the history of research on thinking

Mike Doherty,
Bowling Green University, USA
E-mail: mdoher2@bgnet.bgsu.edu

Ryan Tweney and I have a Brunswik chapter in press for a book edited by Ken Manktelow. It is on Brunswik's contribution to the history and theory of research on thinking. I do not know what the expected publication date is; we have not seen page proofs yet.

I have a chapter in the forthcoming book edited by Sandy Schneider and Jim Shanteau for Cambridge University Press. The book has chapters roaming all over JDM, and my chapter was a commentary on the other contributions. There was no representation of SJT in the book, so I did include a discussion in my chapter. I also called for more attention to situation sampling and to vicarious functioning.

The two presentations at the Brunswik Society will be as follows:

1. The paper by myself, Ryan Tweney & Yanlong Sun will be about the "rule of one variable." It will not be so much a Brunswikian paper as an argument from another perspective that is consistent with Brunswik's reservations about classical experimental design. It presents a formal argument that alternative hypotheses are crucial in null hypothesis testing, and a second argument showing that alternative hypotheses are much more abundant than generally assumed.

2. The paper by John Leach and myself is a report on his dissertation. If he's there, he'll present it. Otherwise I will. Here's the abstract.

Diagnoses and judgments of probability were analyzed in relation to external representation and style of sampling (i.e., completely natural, summarized frequency, and summarized frequency of yoked data). External representation did not produce a reliable effect. All conditions produced a high number of accurate diagnoses. Subjects exposed to completely natural sampling were 100% accurate in diagnosing the disease. However, fewer than 30% of all subjects reported the exact Bayesian probability. A pseudodiagnostic selection effect was obtained under the completely natural and summarized frequency sampling conditions. The disjunction between accuracy on diagnosis and accuracy on the more analytic aspects of this study suggests that different cognitive processes were at work in different parts of the study. The major findings of highly accurate diagnoses and

considerably lower performance on the probability estimate and pseudodiagnostic information selection were consistent with the proposition that the cognitive processes involved in global judgments and diagnoses differ from those involved in analytical reasoning.

Risk of suicide and Decisions to Observe Psychiatric In-Patients

Dawn Dowding, Brodie Paterson, Clare Cassells
University of Stirling
Email: d.w.dowding@stir.ac.uk

We are currently engaged in a study that is examining how psychiatrists and psychiatric nurses in the UK make judgements about patients' suicide risk in an acute in-patient setting, using social judgement approaches. We are particularly interested in exploring how judgements about suicide risk influence decisions about the level of 'observation' ordered in such settings. Case vignettes have been developed, using variables that have been identified from a literature review and consultation with relevant experts. The vignettes are undergoing checks for validity at the minute, and should be distributed to a number of clinicians from acute psychiatric settings within the next 6 weeks. Suicide risk prediction is a complex issue, and data should provide valuable insight into how clinicians use information to inform their judgements.

The use of base rate information as a function of consistency and diagnosticity

Phil Dunwoody
Crawfordville, GA
E-mail: drdunwoody@yahoo.com

Myself, Adam Goodie, and Rob Mahan have recently finished a study examining how experience with base rates influences base rate usage. The ms is presently under review. I'm working as a land surveyor and putting my PhD to good use raising my 10 month old boy.

The use of base rate information has been widely studied in decision making with the conclusion that people underweight base rate information when compared to a normative standard. Three experiments demonstrate that base rate usage under direct experience is moderated by the consistency as well as the diagnosticity of base rate information. Experiment 1 shows that participants use base rate information more when it is consistent than when it is inconsistent. In Experiment 2, this effect was replicated, and transferred to base rate sensitivity in verbal questions posed subsequently. In

Experiment 3, participants' use of base rate information was once again moderated by its consistency, but this effect was itself moderated by the diagnosticity of the base rate information. These studies demonstrate that base rate usage can be an adaptive response to environmental contingencies.

Towards a balanced social psychology

David Funder
University of California, Riverside,
E-mail: funder@citrus.ucr.edu

Social psychology is remarkable for its negativity, in both its behavioral and cognitive research. Behavioral social psychology has emphasized bad behaviors such as obedience to malevolent authority and mindless conformity. Cognitive social psychology, taking its cue from the mainstream study of judgment, emphasizes the errors people make. This emphasis is misguided because it focuses on outcomes rather than processes. Obviously -- to a Brunswikian -- the same process can and probably will lead to both good and bad outcomes depending upon circumstances. Joachim Krueger and I have written a paper -- currently in draft form under editorial review -- addressing this issue and related matters. You can see it at <http://www.psych.ucr.edu/faculty/funder/rap/bbs.pdf>
Comments welcome.

Assessing Clinical Decisions of Medical Students

Bob Galbraith, Tim Siftar, Tony LaDuca and Steve Clyman (National Board of Medical Examiners, Philadelphia, PA)
E-mail: RGalbraith@NBME.org

We are investigating the use of Clinical Judgment Analysis (CJA) together with other approaches to assess decision making around diagnosis and management in third year (M3) medical students during their pediatric clerkship. A project with Pediatric Clerkship Directors from 15-20 schools in the US is being implemented with advice from Tom Stewart, Tom Tape and Paul Sorum. The first step will be to agree upon 8-12 clinical tasks and to accumulate 50 or more actual cases for each. Testing is projected to start in July 2003, and will be accomplished online using a web-based application that provides both analysis (accuracy, cue utilization etc.) and either cognitive or outcome feedback. The study design should allow us to evaluate the effectiveness of CJA for formative as well as summative assessment among medical students.

Applying real-time cognitive feedback in a dental diagnosis task

Gordon Gattie and Ann Bisantz
University at Buffalo,
The State University of New York
E-mail: ggattie@acsu.buffalo.edu and
bisantz@eng.buffalo.edu

Our research examined the effect of ecological predictability and cognitive feedback component on a diagnosis task. We developed a software program requiring participants to make 50 medical judgments of oral diseases, based on an abbreviated case history, color photograph, and cognitive feedback updated after every trial. Therefore, not all cues were explicitly provided to participants. Thirty-six experts-in-training, mostly dental students from the University at Buffalo School of Dental Medicine, and 36 medically naïve participants completed the experiment. Ecological predictability levels were relatively high and low. CFB consisted of the appropriate Lens Model parameters (i.e. ecological predictability was presented in the task information condition) and standardized cue weights. All values were presented numerically.

We found that the Lens Model could be successfully applied to an environment with ill-defined cues, as the experts-in-training performed significantly better than their novice counterparts. In addition, participants in the high ecological predictability condition performed significantly higher than those in the low ecological predictability condition. Experts-in-training were more consistent in their applying their judgment policies than novices.

As with previous CFB studies, participants receiving CFB scored higher than participants receiving outcome feedback. However, we discovered that novices scored higher when receiving cognitive information instead of task information. This results suggests that as people make the transition from novice to expert, they shift from relying on their own decision policies to following actual cue weights determined by the environment. Our results also showed that the performance of all participant groups initially decreased with the introduction of real-time CFB updated after every trial. However, the achievement and cognitive control correlations increased in the latter half of the experiment. Our next step is providing participants with different CFB indices and presenting real-time CFB graphically.

Brunswikian Architecture

Robert Gifford
Victoria BC Canada
E-mail: rgifford@uvic.ca

Reviewing all the newsletters online, I see that our group has only contributed to it 3 times, the last time in 1999, but we have never stopped conducting lens model studies over the last 20 years.

In terms of a lens model paradigm, we appear to be fixed at the place where Brunswik (1945) left off (see the 1956 book, pages 26-29, Experiment D), but this simple paradigm has been employed in contexts ranging from personnel hiring to judged intelligence and personality, burglars' and police assessments of potential burglary targets, and human physical attractiveness.

Our last three published Brunswikian articles have concentrated on how different groups of key players judge building aesthetics. Studies in *Environment and Behavior* (2001) and *Journal of Architectural and Planning Research* (2002) examine how architects and laypersons use different objective building cues (fenestration, cladding, height, shape, etc.) to arrive at different aesthetic conclusions. It has long been known that architects judge buildings differently from the rest of us, but these studies begin to elucidate exactly how and why that happens, beginning with objective cues, and including the observer's intermediate-level assessments such as the building's judged complexity, coherence, friendliness, etc. A similar study (*Educational Research*, 2001) compared student and professor judgements of university classrooms.

The architecture studies seemed to catch the media eye, and were summarized in the national Canadian university newspaper *University Affairs*, the UK popular science journal *NewScientist*, and a few other places.

Are all judgments equal?

Claudia Gonzalez Vallejo
Ohio University
E-mail: gonzalez@ohiou.edu

My Brunswikian research continues to look at the extent to which global judgments differ in systematic ways from judgments performed in a piece by piece fashion (disaggregated-where attribute levels are judged for their quality). In earlier work we found that the relationship between preference orders of objects evaluated via the global method and choice preference orders was smaller than that of choice and disaggregated orders. We are further pursuing this. In addition, we are looking

at whether irrelevant cues influence global judgments to a greater extent than do disaggregated ones in a job selection task.

Using Judgment Analysis to Predict the Targets of Crime

Louise Gunderson
University of Virginia
E-mail: lfq4a@virginia.edu

I am using Judgment Analysis to create a model of target selection by criminals. Generally, when a criminal commits a crime (robbery, burglary) the details of the crime are recorded in police reports. This allows the investigator to collect information about the crime such as location, time, date, and type of offence. Many of the features of the target can be determined from this information, the demographics of the location can be found from the US Census block group information, the weather can be found from the date, and so on. I am using the lens model to construct data mining methodologies to discover the preferences of the criminals.

One of the preliminary results was a methodology to predict the type of object stolen from homes, from a set of objects that were not stolen together in the period of study. First, association rules were used to find the objects that were not stolen together. For the city of Richmond, Virginia in 1996, bicycles, consumable goods, firearms, livestock, and tools were never stolen together. All of the crimes in which one of these objects was stolen were selected. A classification tree was built using the data from 1996-1997, and the features derived from the police reports. This classification tree was used to predict the stolen objects in the crimes occurring in 1998. This method correctly predicted the stolen object 44.2 % of the time, as compared to 28% of the time by random draw.

Further work is continuing on the discovery of target preferences by a clustering methodology based on the lens model. This methodology has been demonstrated to outperform other clustering methods on synthetic data sets. At this point, it is being modified to take into account the distribution of features in the natural environment.

This work has been presented at various engineering conferences, such as PerMIS 2002 and IEEE SMC, 2002. The work presented at last year's Brunswik Society meeting has been published as a chapter in *Advances in Computers*, Vol 56, M. V. Zelkowitz. Ed.

Prostate Cancer Screening: The Lens Model Clarifies a Comparison Between the Health Beliefs Model and a Descriptive Expected Utility Model.

Robert M. Hamm
 Department of Family and Preventive Medicine
 University of Oklahoma Health Sciences Center
robert-hamm@ouhsc.edu

I applied a lens model analysis to cross person data, concerning beliefs about the benefits of prostate cancer screening. The judgment was the man's stated intention to get PSA screening in the next six months. The criterion was his self report, after 6 months, of whether he had gotten PSA screening. For predictors, there were 19 questions, on a 1 to 5 scale. I had also combined the 19 items into theoretically inspired indices: a) 4 Health Beliefs Model measures: benefits, barriers, severity and susceptibility; or b) 10 Descriptive Expected Utility Model component measures; and also c) the EU components were organized through a decision tree into a single measure, the difference between the EU of getting screened and the EU of not getting screened.

The correlation between PSA intention and PSA behavior (each 1 or 0) was .315. The lens model parameters using the 19 items, the 4 HBM measures, the 10 EU components, and the 1 EU difference score, were as shown in this table:

	Linear	Nonlinear
All 19 items	0.117	0.198
4 HBM factors	0.059	0.235
10 EU components	0.072	0.213
EU difference	0.052	0.234

While it appears the models are inadequate, since even the full linear model explains less than half of the correlation, this may be due to the binary nature of the predicted variables. I would like to talk with those of you who have explored logistic regression versions of the lens model at the November Brunswik meeting, or hear from you by email.

More detailed inspection of the Lens Model components shows that Rs (the predictability of the PSA screening intention) was higher than Re (the predictability of the actual behavior), which is plausible since the intention was stated on the same day, while the behavior happened during the subsequent six months. Curious, also, is the perfect correlation between the predictions of the two models, for the one-predictor (EU difference) Lens Model.

	Re	Rs	G	C
All 19 items	0.368	0.475	0.668	0.242
4 HBM factors	0.197	0.296	0.96	0.277
10 EU components	0.312	0.364	0.72	0.263
1 EU difference	0.206	0.275	1	0.274

The exercise of applying the Lens Model gave a different perspective than the initial analyses I did of these data, comparing alternative logistic regression models.

Progress?

K.R.Hammond
krhammond@earthlink.net

I continue to work on the development of a book ms tentatively titled "The Structure of Human Judgment and the Softening of Rationality" but I am having trouble developing the appropriate theme for this book. However, a new event has occurred that all Brunswikians should appreciate, to which I now turn.

Last year Elise Weaver alerted me to the appearance of a book titled "The Number Sense" by a French neuropsychologist (and mathematician), Stanislas Dehaene. It is indeed a splendid book but most important is the fact that it provides neuropsychological support for cognitive continuum theory, and the concept of quasirationality, although Dehaene knew (knows) nothing of either. As a result I wrote a draft chapter in my ms on his work and asked OUP to send it to him for his approval, which he gave. In addition, however, he sent an attachment containing an article he and his colleagues published in Science in 1999, that can only be described as striking. It is a great step forward for Brunswikian theory. The article can be found at <http://www.sciencemag.org/cgi/content/full/284/5416/970> and is reprinted as an addendum to the paper version of the newsletter with permission from the American Association for the Advancement of Science. The full reference for this article is Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999) Sources of Mathematical Thinking: Behavioral and Brain-Imaging Evidence. Science, 284, pp. 970-974 (7th May 1999). Dehaene also has an article on the Edge website at: http://www.edge.org/3rd_culture/dehaene/dehaene_p2.html.

Information search, use, reasoning and being strategic (but never all at once)

Clare Harries,
Centre for Decision Research,
Leeds University Business School
E-mail: ch@lubs.leeds.ac.uk

I've continued work on three or so lines of research (see Harvey, this newsletter). See the Brunswik website www.brunswik.org for a report on the one day international meeting on Clinical Judgment Analysis.

Vicarious functioning, vicarious reasoning and inequalities in health care

With Damien Forrest, Nigel Harvey and Ann Bowling.

In our three year project on the role of decision making in inequalities in the British healthcare system we have extended a classic clinical judgement analysis in several innovative ways and with interesting results. Each physician asked for the information and tests they wanted prior to making a multivariate treatment decision for each of 72 patients. This process focus allows us to include testing for heart disease as both a decision to be analysed, and as a piece of information influencing subsequent treatment decisions. It also allows us to measure the influence of information in relation to the proportion of cases on which it was sought. We elicited subjective policies not as ratings but as a series of graphs indicating the relationship between each level of cue, and the decision (we picked 5 or 6 decisions that we asked them to focus on). Thus we measured the subjective functional relationship between cue and decision and avoided ambiguity of the meaning of ratings. In addition to this, participants listed reasons why they would do each treatment or test and reasons why they would not do that treatment or test.

The preliminary results are striking. Analysis of the idiographic patterns of behaviour show that (only!) half of the physicians in each specialist group (Cardiology, Care of the Elderly, and General Practice) were less likely to treat the old and they did so in terms of different aspects of their patient management (from the amount of basic clinical information they collected, to ordering tests, ordering angiograms, and revascularization). Elderly people with chest pain and suspected angina in the UK are vulnerable to a cumulative effect of discrimination. Most importantly, different physicians gave different reasons for their decision making. For example, some physicians implicated age (in our rationing based society) as a simple contraindication to treatment (a "fair innings" argument), others implicated age as a cue to co-morbidity, which was itself seen as a contraindication to both testing and treatment, or as a cue to reduced potential benefit;

other physicians implied that old age was a cue to a lack of desire for treatment. The end pattern of age-related decision making is a product of both vicarious functioning and vicarious reasoning.

Strategy and knowledge in decision making

I have finally moved from the theoretical analysis of a Brunswikian perspective on strategic decision making to a series of studies that tease out different parts of the process. These studies build on work on dynamic decision making, complex problem solving as well as lens model analysis. One focus is on the strategic decisions that change the ecological validity of predictive cues. Those decisions that change the levels of each cue can be seen as more tactical. Another focus is on the distinction between uncontrollable and controllable aspects of the environment. We'll see where all this gets me next year.

Information search in simple heuristics

Mandeep K. Dhimi, and I continue to work on the collaboration that we started five years, four countries and five cities ago. We compare physicians' information selection to the search specified in simple (matching heuristic) models of their decision making. These simple heuristics are good at predicting physicians' decisions, but of course the physicians select many more pieces of information than specified in the model. For most physicians the simple one-reason for decision making identified by the model is the cue looked at first, sometimes it's the cue looked at last, and for one or two physicians it's the cue looked at after some information and before other information. This work will be presented at the Brunswik meeting this year.

Advice taking and trust in advisors

Nigel Harvey
University College London
n.harvey@ucl.ac.uk

Together with Clare Harries, I have been continuing work on the advice-taking paradigm. In this task, all cues and the criterion refer to the same variable. For example, people are given forecasts of sales of a product that have been produced by four different advisors. On the basis of this information, they make their own best estimate of what sales will be. Some advisors are better than others but judges can find out about this only by experiencing outcome feedback (which informs them of true sales figures for the product). We have directly compared results from this task with those from an equivalent multiple cue probability learning task. In the latter case, product sales had to be forecast from different variables,

such as marketing spend and competitor activity but the underlying formal relation between cues and criterion was the same. We found that the advice-taking task was easier from the start and that learning in it was relatively rapid. Recently, we have extended this paradigm to allow us to investigate the trust that people place in their advisors and factors that affect it.

Brunswikian research at the University of Connecticut

Jim Holzworth
Storrs, Connecticut
Fall, 2002

E-mail: holz@uconnvm.uconn.edu

Research in the Brunswikian tradition continues at the University of Connecticut.

My work with Steven Mellor on judgments concerning labor union issues continues. Two studies were described in the Brunswik 2000 Newsletter. These studies have now been accepted for publication. Dan O'Shea, Steven Mellor, David LaHuis and I (2002, *Journal of Threat Assessment*, 2, 67-84) report on our study concerning contextual and individual influences on the individual's decision to become a replacement worker during a strike. Steven Mellor, Jim Conway and I (*Experimental Psychology*, in press) report on our study concerning people's inclinations to be represented by labor unions.

We continue investigating cognitive continuum theory (CCT). I am continuing my collection of biographical data (biodata) from university students in an attempt to relate it to styles of inductive reasoning. Dennis Thomas and Kathlea Vaughn are investigating a possible relationship between the cognitive continuum index proposed within CCT and Seymour Epstein's rational-experiential index developed within his cognitive-experiential self-theory (CEST). James Pratt and Liz Kramer are investigating cognitive styles and strategies in visual discrimination of computer displayed graphics. Holistic and analytical discrimination judgment strategies are being examined, along with task characteristics, training, and display designs.

Kris Canali, Carrie Nelson, and I continue research comparing methods of judgment analysis. We are comparing the more traditional representative design of judgment tasks with a more efficient representative design suggested by Gary McClelland. As reported last year, we are finding that the efficient representative design accounts for significantly more judgment variance in judgment analysis, and there are no negative consequences on cross-validation of judgment policies.

Applying Lens to Free Flight: A Lens model analysis of pilot and controller decision making in a future ATM system

Pratik D Jha, Ann M Bisantz and Raja Parasuraman*

Department of Industrial Engineering
University at Buffalo,
The State University of New York
Buffalo, New York 14260
E-mail: pjha@eng.buffalo.edu
bisantz@eng.buffalo.edu

*Cognitive Science Laboratory, The Catholic University of America, Washington, DC
parasuraman@cua.edu

Free flight (RTCA, 1995; Parasuraman, Hilburn, & Hoekstra, 2001) and related concepts such as Distributed Air-Ground Traffic Management (DAG-TM) (NASA, 1999) will fundamentally change the authority structure in future air traffic management. A move from management by delegation to management by exception is proposed (Billings, 1996), under which the separation assurance function will be a prime responsibility of pilots. However, controllers will still act in the supervisory role and will have the ultimate responsibility of running the system safely. Research on conflict detection has shown that controllers will be unable to perform this task efficiently under high traffic loads without automation tools such as conflict probes (Galster et al., 2001; Metzger & Parasuraman, 2001). Moreover, initial research on conflict resolution indicates that pilots and controllers have different styles of solving conflict (AGIE) (FAA and NASA, 2002). Controllers solve the conflict earlier than pilots and use more altitude and heading clearances while pilots use more speed and heading changes. Also, it has been noted that there is an apparent mismatch of expectations and biases between the pilots and air traffic controllers in their attempt to solve conflicts. One of the reasons cited was that controllers and pilots have different resolution strategies, due to their fundamentally different working goals. Pilots are more "aircraft centric" as compared to controllers who are likely to focus on an entire chunk of airspace, and therefore act in a more "airspace centric" way when making their decisions.

Based on this premise it is likely that there will be situations of interpersonal conflict between pilots and controllers, in their efforts to resolve conflicts. Decisions may be made using different strategies, as well as at different times. Modeling these judgment differences may provide the

means both to understand these conflicts more precisely, and to provide indications regarding necessary training or real-time displays and feedback which could mitigate conflicts.

A candidate model for describing and comparing differences between pilot and controller differences is the Lens Model (Brunswik, 1955; Cooksey, 1996). The Lens Model, as typically instantiated, provides dual, symmetric models of both the human judge and the environment and can be used for capturing and comparing decision policies among different players in the free flight environment. The Lens Model has been successfully applied in numerous contexts (see Cooksey, 1996) such as medical decision-making and social policy judgments, to describe aspects of judgment performance. It has also been used in the past for analyzing judgment which are related to situations of interpersonal conflict (Brehmer, 1976). Recently efforts have been made to apply it into more dynamic environments like command and control (Bisantz et al., 2000), fault categorization and diagnosis in process control (Jha, 2001) and aircraft collision avoidance (Pritchett and Bisantz, 2002). In particular, in the latter study, the Lens Model was used to compare judgments made by different automated and human agents, using parameters from the Lens model to characterize differences among the judgment policies.

We are beginning a new research effort to examine the appropriateness of Lens Models of judgment for describing and supporting the resolution of judgment conflicts in the proposed new air traffic management concept called free flight.

Exemplar matching and cue criterion relations

Peter Juslin
Umeå University
E-mail: peter.juslin@psy.umu.se

We are pursuing our research that attempts to relate traditional multiple cue probability learning to basic research on cognitive processes and representations in cognitive science. A first study suggested that people do integrate multiple cues in multiple cue judgment tasks, either implicitly by retrieving similar exemplars, or explicitly by abstracting cue-criterion relations during training that are later integrated into a judgment. In one series of experiments we have demonstrated that when the criterion of judgment is changed from a binary to a continuous variable (e.g., from a binary categorization task to a standard multiple cue judgment task) the processes shift from primarily exemplar-based processes to processes that involve abstraction of explicit cue-criterion relations. A couple of experiments, which we are just in the process of writing up, show that when the cue

combination rule is additive people are prone to rely on explicit cue abstraction, but when the combination rule is multiplicative, they turn to exemplar memory. These studies are in various stages of preparation or revision and should hopefully appear in print in the not so distant future. An underlying ambition with this line of research is to attain a more detailed cognitive understanding of the notions of "intuition" and "analysis", as well as to relate and interpret the lens model components in terms of cognitive processes.

Another project aims to understand what sort of knowledge people use to make probability judgments -- for example, exemplar knowledge, prototype similarity (representativeness), relative frequency, or frequency -- by modeling participants' probability judgments in controlled laboratory learning experiments. The background is that these theories are routinely invoked in various contexts and used to interpret findings post hoc, or for motivating general predictions about presence or absence of bias (e.g., overconfidence), but they are rarely tested (or even specified) in a more rigorous way. (And this goes for my own research too.) The purpose is to understand if and when people use these sorts of knowledge to make probability judgments as a function of the judge, the specific task, and the ecology.

Tracing shoe tracks

José Kerstholt
TNO Human Factors
Email: kerstholt@tm.tno.nl

Pieter Koele
Universiteit van Amsterdam
Email: ml_koele@macmail.psy.uva.nl

This month, Roos Paashuis started her research on how forensic researchers make their judgements. Verifiability is of utmost importance in forensic research. However, much of this research has a non-routine character, raising the question as to what extent interpretations and judgements can be made transparent. Roos will initially focus on one particular type of forensic research, the comparison of shoe tracks found at a crime scene with a particular shoe. The major goal of the PhD project is to characterise the process underlying these comparative judgements for both experts and novices, using Policy Capturing techniques. There seem to be two separate phases in this process: one in which the researcher identifies which cues are worth considering and a second one in which track and shoe are compared on these cues. As forensic researchers are supposed to explain their judgements in court (mostly in writing), the relation between the actual process

and the subjective description will be addressed as well. In a number of cases, forensic researchers also provide a certainty score to their judgement. Another aim of the project is therefore to examine the interpretation and quantification of uncertainty by forensic researchers.

Toward a Coherent Ecological Psychology

Alex Kirlik,
University of Illinois at Urbana-Champaign
E-mail: kirlik@uiuc.edu

I spent the 2001-2002 academic year visiting Haskins Laboratories, New Haven, CT, the PACE Center (Psychology of Abilities, Competencies and Expertise) at Yale University, and the Department of Psychology at the University of Connecticut (gracefully received by Jim Holzworth and the rest of the I/O Program, and also by the Center for the Ecological Study of Perception and Action). After my initial interactions with scholars such as Holzworth, Turvey, Sternberg, Mace, Michaels, Shaw, etc., (and especially their students), I couldn't help but to believe all these theorists agreed on fundamental issues, but that historical accidents in the development of ecological theory had caused fractures, perhaps self-serving in the short run, but truly destructive in the longer term. I thus assigned myself the task of doing the detective work of trying to understand the reasons underlying the divide between the two most prominent schools of ecological psychology (Brunswikian, Gibsonian), and how I might present my findings in a way that was perceived as constructive by members of both intellectual communities. My year culminated with a presentation at Haskins Laboratories, in a talk with the following abstract (the full Powerpoint slides can be downloaded from here <http://homepage.mac.com/alexkirlik/FileSharing1.html>).

"The selection of basic research questions is driven by judgments of what it is that makes research fundamental. Whatever else (elegance, etc.) might be meant by this idea, generalizability is a key, and many would say, defining, element. Fundamental research findings should generalize beyond the context of discovery, either to a specified domain, or in the ideal, apply universally. I describe how, unlike scientists in many other fields, experimental psychologists successfully use at least three different logics for generalizing findings beyond the context of discovery. A failure to recognize the legitimacy of all three engines of generalization, and to distinguish between them when required, lends confusion to both theoretical and methodological discussions. I illustrate this thesis by discussing, and attempting to resolve, the tremendous amount of confusion that has grown up around the concept of "ecological validity." This work is a product of

working closely with scholars from both the Brunswikian and Gibsonian ecological traditions for ten years, and my attempt to view these traditions, among others, as complementary rather than as conflicting. While this work was motivated by a desire to create a coherent ecological psychology, I conclude by discussing the methodological implications of this work for the broader domain of psychological science, broadly conceived."

Configural judgement strategies and MCPL

Pieter Koele
Universiteit van Amsterdam
Email: ml_koele@macmail.psy.uva.nl

Mick de Niet (an undergraduate student of mine) has just completed an MCPL experiment in which the judgement task consisted of predicting the severity of Brunswik's Disease of fictitious patients on the basis of two labelled cues, reflecting the concentration of certain substances in the blood of the patients. OF was provided after each trial, and subjects had to judge 90 patients. In condition I (n = 46) the cues (X1 and X2) and the criterion Ye were related thus:

$$\begin{aligned} Y_e &= 0.125 X_1 + 9.2 && \text{if } X_2 < \text{Mean } X_2 \\ Y_e &= X_1 && \text{if } X_2 \geq \text{Mean } X_2 \end{aligned}$$

In condition II (also n = 46) the relations were:

$$\begin{aligned} Y_e &= -X_1 + 21 && \text{if } X_2 < \text{Mean } X_2 \\ Y_e &= X_1 && \text{if } X_2 \geq \text{Mean } X_2 \end{aligned}$$

X1 and X2 values were randomly drawn from a uniform [1; 20] distribution. So, in both conditions the criterion was perfectly predictable, but not by consistently using a linear additive combination of the two cues. In condition I the best fitting linear combination of X1 and X2 yields a multiple correlation of 0.82, in condition II it is only 0.12.

Analyses of the experimental results (based on dividing the 90 trials into three blocks of 30 trials each, and calculating the lens model measures in each block) showed that in condition I many subjects managed to find a strategy similar to the best fitting linear additive strategy (mean achievement in block 3 is 0.74, mean Rs = 0.88, mean G = 0.99). Configural cue use did not reach a substantial level (mean C = .24 in block 3). In condition II subject overall achievement was low in block 3 (mean achievement = 0.36, mean Rs = 0.52, mean G = 0.05), but substantially better than that of the best fitting linear model. And indeed, this was the result of configural cue use: mean C = 0.42 in block 3. We think this shows that subjects are able to detect to a certain extent configurality in

an MCPL task, but only when improper linear models fail to yield (nearly) correct predictions. And this in spite of the sub optimality of Outcome Feedback.

Following the learning task we gave the subjects in both conditions another block of 30 trials, now without OF, but after providing them with precise information about the task structure. In condition I subjects did slightly better than in block 3 of the learning phase (mean achievement = 0.82) and this was mainly the result of an increase in configural cue use: mean C = 0.36. In condition II mean achievement increased to 0.53, but this was mainly the result of an increase in linear task knowledge: mean G = 0.67, whereas mean C = 0.57.

As I am writing this abstract we are still a bit puzzled about these last results. It might be conceivable that a certain amount of cognitive laziness restrained the subjects in condition I from switching from a rather successful improper strategy to a configural strategy, but we expected the subjects in condition II to do much better. Or does this demonstrate once again that people find it very difficult to work with negative linear relations? I'll start working on a serious manuscript about this experiment shortly.

Different kinds of confidence

Josh Klayman
University of Chicago
E-mail: joshk@uchicago.edu

I'm continuing my interest in subjective confidence, looking at different kinds thereof. Jack Soll (INSEAD) and I are studying confidence expressed in terms of subjective intervals ("I'm 80% sure that the planet Neptune was discovered between 1760 and 1940"...). We have two main findings of interest.

1. Even using random sampling from whole domains (to approximate representative sampling), subjective intervals show substantial overconfidence. That is, one's 80% intervals contain much less than 80% of the correct answers. Some of this is due to biasing effects of unbiased error in setting the endpoints of the interval. However, we demonstrate that subjective intervals are also systematically too narrow.

2. How much too narrow they are depends on exactly how you ask for them. If, as above, you ask for a single interval, the intervals are barely half the size that a well-calibrated person would need. If you ask instead for two judgments--"I'm 90% sure it's greater than ... and I'm 90% sure it's less than ...", the implied 80% interval is larger--about 2/3 of the well-calibrated size. If you also ask for a median

estimate along with the two ends ("It's about equally likely to be more or less than ..."), then the interval is yet larger, almost the well-calibrated size.

Why? Well... we're working on that. We'd be very interested to hear the hypotheses of our fellow Brunswikians!

The second kind of confidence I'm working on has to do with judgments about where one's performance stands relative to others'. In work with Katherine Burson (here at Chicago) and Rick Larrick (Duke U.), we first weigh in on a recent debate. It seems that the poorest performers on a task are also those who most overestimate their standing relative to others. Is this an interesting cognitive phenomenon (a correlation between cognitive ability and the metacognitive ability to know how skilled you are) or the result of noisy estimates, leading to regression toward the mean, plus an overall upward bias? We find for the latter, by testing tasks in which there is no overall upward bias. Plus, we replicate earlier findings that people are positively biased in judgments of where they stand relative to others only when they find the task to be easy. When the task is difficult, they are actually negatively biased. So, now, we're looking to create a new debate: Does this represent an egocentric anchoring on one's own ease of performance, with insufficient adjustment for what one knows about others' ability to do the task? Or is it really a manifestation of people's use of two probabilistic cues to their standing, namely their impression of how hard it was for them and their impression of how hard it would be for other people? We're still working on that, but you can probably guess that a member of this society would tend toward the latter explanation... Not that we would let prior beliefs color our interpretation of the data (not more than would be normatively appropriate, anyway!)

More on how he developed his ideas

Elke Kurz-Milcke,
College of Computing,
Georgia Institute of Technology
E-mail: kurzmi@cc.gatech.edu

Volume 5 of a series titled *Portraits of Pioneers in Psychology* edited by G. A. Kimble and M. Wertheimer (published by Erlbaum) is soon to appear and will contain a chapter on Brunswik. The chapter is titled *Egon Brunswik: Student of achievement* and was written in a transatlantic collaboration between Nancy Innis from the University of Western Ontario and myself, then working at the University of Tübingen in Germany. The chapter gives an overview to Brunswik's life and work, emphasizing the period of time just after he arrived in the United States, a period of

transition. The chapter describes the ways in which Edward Chace Tolman played an active part in Brunswik's transition from Vienna to Berkeley.

In a way, the development of the lens analogy into the lens model brackets this transition period: Still in Vienna, Brunswik designed a geometrical lens analogy (using only straight lines), at the end of the transition stood an icon (showing mostly curved lines). Initially, developed as a graphical representation of Bühler's duplicity principle, Brunswik subsequently used the lens analogy to communicate his approach to perceptual achievement (alias constancy) to an English-speaking audience. With the development of his program of probabilistic functionalism, vicarious functioning became a predominant principle in Brunswik's account of organismic achievement. Correspondingly, the lens model was not only informed by the lens analogy but also by other metaphors, such as "hierarchy" and "arch," which are especially notable in Brunswik's elaborate *Figure 10* at the close of Part I of *Perception and the Representative Design of Psychological Experiments*. Only later, after its original publication in 1947, did Brunswik refer to this particular figure as "lens-like." Thus, the lens model was developed by bringing additional metaphors on board; it was developed as a mixed metaphor and has been such ever since.

In an effort to find a suitable photograph of Brunswik, which is customary in this series of *Portraits*, we contacted various institutions in Austria and also the Gazi Institute in Turkey, where Brunswik had spent 1931-1932 as a visiting lecture. Thanks to a number of Turkish colleagues we were able to obtain the photograph of Brunswik showing him with colleagues and graduate students in Ankara, which is available for your inspection at the Brunswik Society's website. Brunswik looks rather formal in this picture and in my humble judgement was not yet the person who would design the lens model icon in the 1955 article in *Psychological Review* with its variably curved connections between cues, variable, and response, and with the underlying "laissez-faire policy for the ecology."

Representative design and social perception

Clark Leavitt
Ohio State University
E-mail: leavitt.2@osu.edu

I am currently writing up the first study in a program designed to explore the ramifications of applying ecologically representative design to responses used in social perception research. The study briefly reviews the roots of the ecological approach in the work of both Brunswik and Gibson, pointing to the relative neglect of procedures for formal definition

and selection of response variables in both traditions. Recent research in social perception using the 'ecological approach', particularly the perception of traits from reduced stimuli, is examined and commonly used procedures are indicated. Factor analyses of ratings of snapshots are presented to demonstrate the effects of neglecting response selection procedures. Dimensions revealed by the factor structure reflect basic adaptive behavior.

The second study will apply representative design of response variables to ratings of ordinary daily events to test the emergence of the hypothesized adaptive dimensions.

Clinical Judgment Analysis of Patient Prioritisation for Elective General Surgery

Andrew D. MacCormick
Faculty of Medical and Health Sciences,
University of Auckland
E-mail: andrew.maccormick@auckland.ac.nz

Background

The basic topic of investigation is the prioritisation of patients for elective general surgery. In New Zealand the health care system is state funded. There is excess demand over supply of elective surgical services. Previously clinicians had organised waiting lists for operations on the basis of three categories, namely; urgent, semi urgent and routine. In reforms undertaken in the 1990's there was an attempt to move away from waiting lists to a system by which patients were given certainty as to if and when they would receive surgery (the "booking system"). To do this Clinical Priority Assessment Criteria (CPAC) were developed. These are algorithms by which a score is generated. This score is used to determine a patient's access to surgery. The benefit of the algorithms was that they provided transparency to the prioritisation process and could possibly lead to improved reliability. Unfortunately some of these CPAC were designed without clinical input and didn't reflect clinical judgement. For example one algorithm failed to prioritise patients with malignant disease.

Therefore we felt that SJT was applicable. It would enable us to provide a transparent method of prioritising that still reflected clinical judgement.

Work to date

Our first step was to determine the cues/criteria that surgeons felt they used in priority setting. This was performed using the repertory grid technique.

We then investigated how surgeons used these criteria i.e. their policy. This was performed using clinical vignettes that were computer

administered to sixty surgeons. When it came to prioritising the vignettes rather than allowing the surgeons to see the assigned level of each cue we allowed them to subjectively interpret the cue level using a Visual Analogue Scale (0-50mm) for each cue. They also determined their global judgement of priority using a VAS (0-100mm).

We then undertook multiple regression on an idiographic basis using the global judgement of priority as the dependent variable and the seven cues as independent variables. We then performed cluster analysis of the individual surgeons to determine if there were any different philosophical groupings to how the criteria were used. This revealed two groups. We also undertook a nomothetic analysis to determine how the global functioning of such a method of priority scoring might be applied to our specific situation.

Future work

Given that there were two different philosophical groupings of cue use we are interested to see if we can develop consensus between these two groups. As such we are undertaking a second round of the sixty surgeons assessment of priority of the vignettes. We would like to use feedback/feedforward using the information derived from the first round.

Secondly we are undertaking a cross-validation of the approach using real patients in a clinical setting. As an extension of this we would like to perform real time feedback to surgeons in the use of the criteria/cues. This would be a monthly meeting to feedback to the policy of individual surgeons in our institution.

Coherence and Correspondence in an Automated Flight deck

Kathleen L Mosier
San Francisco State University
E-mail: kmosier@sfsu.edu

This year we have been focusing on the concepts of coherence and correspondence as they define and impact decision-making in the automated cockpit. In a preliminary study, Cathy Jacobson did as her master's thesis an analysis of ASRS (Aviation Safety Reporting System) reports. Data from voluntary incident reports were gathered to examine pilots' use of coherence- and correspondence-based decision-making processes as a function of phase of flight, weather conditions, type of event, and level of aircraft automation. Data were analyzed using Chi Square statistical tests. The findings suggest that pilots' use of either a coherence- or correspondence-based decision strategy is dependent on the situational conditions and

constraints present, namely event type, weather conditions, and phase of flight.

A second study, which we will be conducting on an internet-based platform, focuses on regional airlines and flight decks common to their operations, as this is where we think we may be able to make the most impact regarding automation use. The intent of this study is to identify current practices regarding the use of information toward coherence in decision making, and the impact of operational factors on this process. We are tracking pilot diagnosis and decision making strategies for different types of problems as a function of three important operational variables: a) source of the initial indication of a problem (that is, automated vs non-automated), b) whether or not all information sources are consistent with a particular diagnosis or decision (that is, congruence vs inconsistency of available information), and c) time pressure.

Pilots will respond to a series of scenarios that vary in source of initial information, consistency of information sources, and time pressure, by accessing relevant information until they can make a diagnosis and come to a decision about what to do. The scenarios reflect some of the most common problems cited in incident reports as discussed above, as well as others involving a degree of ambiguity and information search. For each scenario, they will be given one initial piece of information that indicates a problem, and will be asked to access other relevant information by clicking on the instrument or information icon (e.g., for a flight manual). The order and type of information accessed will be tracked so that we can look at information search patterns as a function of the independent variables. Pilots will also be asked to report the level of confidence they have in their diagnoses and decisions. Results of the study will enable us to identify effects of the independent variables on coherent decision processes.

Based on previously discussed work on coherence in automation use and automation bias, it is hypothesized that when pilots see automated information or warnings prior to other information, they will conduct a less thorough and complete information search for diagnosis and decision making than when the initial information comes from a non-automated source. They will also be less likely to identify and resolve inconsistencies in available information than when the initial information comes from a non-automated source. Time pressure is expected to exacerbate the tendency to rely heavily on automated information at the expense of coherence.

A Validation Analysis of Methods of Utility Estimation

Radhika Nath
Rockefeller College,
State University of New York at Albany
E-mail: radicalrad@earthlink.net

For my Ph.D dissertation I examined the validity of three different utility elicitation methods used in medical decision making. These instruments (Visual analog scale, Time tradeoff and Standard gamble) are used to elicit patient preferences which in turn inform physician decision making. Where most previous research looked at preferences for a single health state, in this study the multiple methods were used to estimate preferences for each of three different health states (blindness, stroke and AIDS). This facilitated a thorough MultiTrait MultiMethod evaluation of the validity of the methods. While questions about the instruments have been raised before, I believe this is the first study to undertake a thorough MultiTrait MultiMethod examination of the instruments. I found that the three specific instruments I chose to study did not achieve validity. It is not clear, what the "utilities" achieved through these methods represent and how they relate to each other. In other words, the results obtained from using these methods should be used with caution if at all, in medical decision making and other fields. Further, I also examined if specific individual characteristics (such as a facility with numbers) had any affect on responses to specific methods as hypothesized by other scholars. Using a structural equation model I found that, of the various characteristics studied, only those related to risk (specifically physical and professional risk) had any affect on responses to the time tradeoff and the standard gamble instruments.

Predicting search strategies in simple heuristics and using feedback in MCPL

Ben Newell and David Shanks
University College London
E-mail: b.newell@ucl.ac.uk

Continuing the program of research reported in last year's newsletter David Shanks and I have been furthering our empirical investigations of fast-and-frugal heuristics, with a particular focus on search strategies. In work with Tim Rakow of the University of Essex, we have been examining how decision-makers learn about cue properties in environments with objective criteria. In such environments, cues can be evaluated on the basis of three properties: validity (the probability that a cue identifies the correct choice if cue values differ between alternatives); discrimination rate (the proportion of

occasions on which a cue has differing values); and success (the expected proportion of correct choices when only that cue can be used).

It turns out that validity and discrimination rate are often negatively correlated. For example, in the German cities task (identify the city from a pair that has the largest population) the 'capital city cue' (knowing if either city in the pair is the capital) has a very high validity, because normally a capital has a very large population, but a very low discrimination rate because there is only one capital. This inverse relationship is potentially problematic for heuristics such as "Take-The-Best" (Gigerenzer & Goldstein, 1996), which use a validity-ordered search rule. This because the first few cues looked up by the heuristic, though highly valid will often not discriminate between alternatives and therefore not be able to be used as a basis for a decision. In environments where information is costly such a non-frugal search rule appears maladaptive. Using a stock market prediction task (see Newell & Shanks, 2003) we tested our hypothesis that people would learn a sensitivity to both the validity and the discrimination rate of cues and that their search patterns would follow a pattern that was a function of these two properties - namely success. In two experiments we found support for this hypothesis. The patterns of search through the cues were more closely associated with the search ordering predicted by success than by discrimination rate or validity. Furthermore, when asked to rate each cue for 'usefulness' at the end of the experiment these ratings again were more closely associated with the success ordering than the other two predicted orderings. These results provide strong support for the claim that search through cues in environments with objective criteria is determined by the success rate of cues, and not by validity as the search rule of the "Take-the-Best" heuristic states.

In a new project, unrelated to the fast-and-frugal heuristics, we have been examining the role of feedback in discrete cue MCPL. Using a four cue environment we compared performance of a group given task information feedback with a group given only outcome feedback. The findings mirrored those seen in continuous cue MCPL with the group given task information feedback performing better than the group given only outcome feedback. However, in a follow-up experiment we changed the task structure such that the cues and criterion were presented on the same dimension. Now the difference in performance between the feedback groups largely disappeared due to improvements in the outcome feedback only group. The results are consistent with Harries and Harvey's (2000) contention that having cues and criterion on the same dimension provides participants directly with information

about whether their reliance on particular cues is appropriate or inappropriate – information which is much harder to obtain when cues and criterion are on different dimensions. The results also suggest that all forms of feedback could be effective in MCPL, provided information is presented in a form that allows people to understand the relation between the cues and the outcome.

All reported work is part of the programme of the ESRC Centre for Economic Learning and Social Evolution (ELSE).

Monitoring and aiding clinical decision making under the new Dutch Youth Care Act: Call for suggestions

Huub Pijnenburg
Nijmegen University

Praktikon - centre for research and development in youth care

E-mail: h.pijnenburg@acsw.kun.nl

As a result of the impending new Dutch Youth Care Act (which will most likely come into effect in 2003) a fundamental legal and logistical restructuring of youth care is currently under way in the Netherlands. In the old care structure, separate organisations within essentially independent health, legal and welfare systems sequentially assessed a client's needs in relation to the services that they themselves provided. This amounted to a form of supply side economics: treatment decisions were often not so much the result of diagnostic decision making per se, but rather the result of a matching process. Diagnosticians (employed by organisations for care provision) matched a client's needs against their employer's fixed set of treatment options. This probably exacerbated the identified problems with clinical decision making whereby there was a lack of attention to clients' desires, limited and confirmatory diagnoses, and a lack of relationship between clinical needs and treatment regime. Within such a structure, we found handheld Bayesian based decision aids and diagnostic expert systems to have limited impact on improving clinical decision making. Presumably judgment analysis and feedback of task information would not have fared better. The new system is an attempt to change youth care's supply side economics into a form of 'demand/need side' economics. Its ambitious objectives are: to make youth care more client-oriented, to make diagnostic decision making more explicit and rational, to provide care as close to the client's home as possible, to change the dominant role of care providers, and to break down barriers between care providers within the health, welfare and legal subsystems. Clients now gain access to care via a single 'front door': the local/regional Youth Care Bureau. Here each client's needs are

assessed and treatment decisions are made, based on the findings and conclusions of independent YCB diagnosticians (mostly operating in multidisciplinary teams). Next, treatment is provided accordingly (based on standardised protocols) by an allocated care centre. No longer can care providers turn down clients. As a result we anticipate the quality of clinical decision making to improve. At the same time it is clear that this new set-up is by no means a panacea. Yes, self-interest of care centres will be avoided and diagnostic decision making should be improved, avoiding hypothesis confirmation et cetera. But unless specific measures are put into place, outcome feedback from this two tier system may be just as elusive as it was in the old system. It is clear that for Dutch youth care important times lies ahead. That is why at the moment we are rethinking our R&D priorities at Praktikon and formulating a project agenda for the coming years. We welcome contributions from fellow decision researchers in this process. At the first Clinical Judgement Analysis Meeting, staged earlier this year at Leeds University Business School by Clare Harries, the main recommendation from the audience was that we focus not on analysing individuals' models of decision making, but instead that we should focus on task analysis based on multiple methods including focus groups. In the context of youth care in the Netherlands, as in many clinical contexts, the change of practice and role necessitated by changing circumstances means that there will now be no expert practitioners. As in many other clinical contexts, analysis of the task, and of the multi-faceted environment is therefore of primary importance. My question to the readers of this Newsletter, is: do you agree with this recommendation? What do you believe are key questions we should address? What are suitable and effective research approaches and tools, that will help us realise our ambition to monitor and support this reorganisation and development process, and come up with results that will be relevant to practitioners and researchers alike? Please mail any comments or suggestions you care to make to: h.pijnenburg@acsw.kun.nl

The importance the task environment in processing threatening stimuli

Manuel-Miguel Ramos Álvarez
E-mail: mramos@ujaen.es

Two years ago I entered a research group interested in finding out about cognitive biases and emotional disorders. This group was given a grant for investigating within this field. The research I am going to talk about is connected to this context. A

very interesting result in cognitive biases in connection with emotional disorders is the fact that high anxiety people show a processing bias (on the attention process) when faced with threatening content information, whereas normal anxiety people do not show this bias. To arrive at this conclusion the usual task is to compare the effect provoked by a prior stimulus (prime) on the objective (target), and the attention bias is inferred comparing the stimulation with threatening content (i.e. images or words that trigger anxiety) with a neutral stimulation. Throughout the years, research in this field has introduced the stimulation in a standard way, using a design with neutral covariation between the probe and threat stimuli; that is to say, all the stimuli combinations have the same probability of appearance in the experimental task. However, this kind of design is weak, as it does not consider a very important side, which has to do with the ecological validity of the cues. Given my interest on the lens-model Brunswik approach, I have devoted most of my time to designing a hybrid task that allows considering the cues validity within the experimental context of attention biases (i.e. priming tasks). The experiments we have carried out using this hybrid task show the relevance of the ecological validity in two ways. When there is a very high relation between the cues and the objective stimulus, the results of the attention bias are different than when there is a null relation. Besides, the processing bias has to do with the biological preparation that comes before the differences between anxious and non-anxious people. In fact, we think that this approach could be able to explain the fact that some researchers have found a very weak magnitude in the bias, while others did not even find it.

Decisions to Prescribe Estrogen Replacement Therapy

Michael Ravitch
Northwestern University
E-mail: ravitch@northwestern.edu

Together with Arthur Elstein, David Rovner, Margaret Holmes-Rovner, Gerald Holzman, and Marilyn Rothert we conducted a variety of studies regarding physicians' judgements and decisions in estrogen replacement therapy. We used vignettes to represent the various combinations of the design - cancer risk, fracture risk, severity of hot flashes. These studies were carried out at Michigan State University - Arthur Elstein was the PI.

Rapid Clinical Decision in Context: A theoretical framework to understand physician decision-making

Joshua H Sarver, Neal V Dawson, Susan W Hinze,
Rita K Cydulka, Robert S Wigton, Said A Ibrahim,
David W Baker
Case Western Reserve University, Ohio.
E-mail: sarver@po.cwru.edu

The purpose of this paper is to draw on previous work in multiple disciplines to establish a theoretical framework for clinical decision-making that incorporates non-medical factors, such as race, into the way physicians may make decisions in the real world practice of medicine. The proposed Rapid Clinical Decision in Context (RCDC) model attempts to understand the influence of various contextual elements on physicians' decision-making process. A thorough review of the available literature provides ample support for the RCDC model. In brief, physicians' use of multiple fallible indicators leads to errors in judgment. The use of a recognition-primed decision-making strategy means that physicians are likely to rely on their initial impression of the patient. When physicians and patients interact, the interaction often fails to alter physicians' initial impressions because of communication difficulties arising from social differences between physicians and patients, such as cultural capital and predispositions. Understanding real-world physician decision-making has broad healthcare implications from addressing racial/ethnic disparities in treatment to medical education.

Operators' Judgment Performance with Automated Decision Aids: A theoretical model and empirical investigation

Younho Seong yseong@ncat.edu,
Department of Industrial & Systems Engineering,
North Carolina A&T State University, NC

Ann M. Bisantz bisantz@eng.buffalo.edu,
Department of Industrial Engineering,
State University of New York at Buffalo, NY

In an effort to model the human operators' judgment performance with automated decision aids (ADA), the Lens model with its extension was used. However, the Lens model designs including n-system and hierarchical system design were not suitable to model the environment. The major reason is that a model of human operator with automated decision aid needs to reflect the situation where the operator is provided the "raw" or sensed information along with an estimate from an ADA. Here, an ADA receives the same

information as the operator does and fuses the information to produce an environmental estimate in more comprehensible form, i.e., probability of attack. Therefore, a Lens model was developed, called hybrid Lens model, by combining a triple system and the hierarchical system designs. Associated Lens model parameters were redefined to reflect a variety of characteristics of the judgment agents; human operator, ADA or human judgment with ADA.

In this environment described above, human operators' understanding about the ADA is critical in making decisions whether to accept or reject the decision aid's estimates. This leads us to another important issue in human machine interaction: human operators' trust in ADA. To support human judgment performance and/or calibration of their trust in ADA, the cognitive feedback was used to provide some information about the ADA. The cognitive information feedback of the ADA was provided to increase human operators' understanding of the inner workings of the ADA.

Empirical results showed that participants were significantly affected by the validity of the decision aid's estimates (Re_ADA). More importantly, those provided with the feedback information outperformed participants only with the decision aid in all Lens model parameters. Also, participants were able to utilize the feedback information to calibrate their trust in the decision aid in finer detail.

In conclusion, this study investigated the effect of automated decision aid's various characteristics on operators' judgment performance and trust in the system. A Lens model of human judgment performance with the decision aid was developed to identify the precise effect of ADA on judgment performance. Additionally, the cognitive information feedback was utilized as a way to increase judges' understanding of the decision aid. Results showed that the cognitive information feedback of the decision aid was useful in supporting judges' judgment performance and calibrating their trust in the decision aid. Further, the effect became greater as the "quality" of the decision aid became worse.

Can Decision Errors Be Predicted Before They Happen? Application of a New Measure of Skilled Performance

James Shanteau, along with Rickey Thomas, Brian Friel, John Raacke, Shawn Farris (all at Kansas State University, USA), David Weiss (California State University, Los Angeles, USA), and Julia Pounds (FAA, Oklahoma City, USA)

E-mail: shanteau@ksu.edu

We conducted a two-month longitudinal study of the development of expertise in a simulated air-traffic-control microworld task. The simulation (CTEAM developed by the FAA) requires continuous, dynamic control of multiple aircraft through an air sector. To analyze results, we applied a new moving-window measure of skilled performance. The measure (C-CWS for Continuous-CWS) is an extension of CWS (Cochran-Weiss-Shanteau), developed originally for static analysis of single-case judgments. Our goal was to track the acquisition of expertise over an extended period.

Method

Twelve operators were trained for eight weeks in a single-sector version of CTEAM. Six scenarios were generated by crossing three levels of aircraft density with two levels of airspace restrictions. Each scenario lasts 5 to 8 min. To allow for comparisons, the same 12 aircraft were embedded in all scenarios. The operator's task was to route all aircraft from their origin to their destination. The dependent measures were number of control actions and time to move planes through the sector; similar results were obtained with the two measures.

Results

C-CWS was sensitive to variations in performance associated with both aircraft density and restrictions in airspace. C-CWS was moderately correlated with two "objective measures" of performance - number of separation errors and number of barrier incursions. However, C-CWS outperformed these objective measures in two respects. First, C-CWS revealed performance improvements even after objective measures reached asymptote. Second, C-CWS showed increased sensitivity over sessions, whereas the objective measures became less sensitive as skill improved.

Additional analyses revealed that skill development was closely tied to internal consistency and discrimination (which make up C-CWS). Thus, the components of C-CWS were validated. However, skill development was only marginally related to consensus (which is not part of C-CWS). Most other approaches to assessing

expertise, including correlational measures, rely in part or in all on consensus.

C-CWS provided performance assessments at roughly 1 to 2 min intervals. Unexpectedly, drops in C-CWS values were discovered to be predictive of some (but not all) errors before they occurred. That is, a rapid decline in C-CWS was often followed by an operational error 1 to 3 min later. It appears that C-CWS may be sensitive to performance deterioration prior to an overt error.

Conclusions

C-CWS proved successful in assessing development of expertise in a complex, dynamically evolving task. Other approaches to assessing expertise were less useful. These findings have implications for the selection, training, and evaluation of experts. In addition, this line of research has consequences for rule learning, development of mental models, influence of context on decision making, and the role of individual differences on decision behavior.

The discovery that falloffs in C-CWS scores may be predictive of subsequent errors has important ramifications. This may provide a performance-based approach to identifying and preventing errors before they occur. We are now exploring the generality of this finding in other contexts.

For more information on CWS, see our website www.ksu.edu/psych/cws

Diagnosing Chronic Heart Failure in Primary Health Care

Ylva Skånér

Karolinska Institute, Stockholm, Sweden
E-mail: ylva.skaneer@klinvet.ki.se

The aim of my project is to describe General Practitioners' (GPs') diagnostic processes and strategies regarding patients with suspected heart failure. I have done two studies with Clinical Judgement Analysis technique (CJA) and I am now supplementing them with a study in which I use Think Aloud technique (TA).

In the two CJA studies I have used case vignettes based on authentic patients. In the first study, the patients were collected from two health centres. In order to get patients with more valid diagnoses (heart-failure or not), the patients in the second study were collected among patients referred from GPs to a cardiology out-patient clinic. GPs, cardiologists and medical students were compared as regards diagnostic accomplishment and diagnostic strategies, and I found them to be similar on the group level, but very different at the individual level. The most important cues for the participants were cardiac enlargement and pulmonary stasis.

Strategies, in which cardiac enlargement was the predominating cue, led to a higher diagnostic accomplishment; a third of the participants used such strategies. The cues given in the vignettes could have been utilized more efficiently; cardiac enlargement seems to be more important while "classical" symptoms are less important for predicting heart failure than the participants themselves realized.

To further analyse the diagnostic process, I will now use TA technique. GPs will be presented with six case descriptions, based on authentic patients from primary health care (a sample of patients utilized in the second study, representing different levels of difficulty). Each case description will be presented on consecutive computer screens (history, symptoms and signs; laboratory and electrocardiogram; X-ray; echocardiography), and the think aloud session will be taped, transcribed, coded, and analysed. Our central questions concern how information about the patient is collected and integrated and how decision rules and knowledge are applied in the decision.

General Practitioners' decision making and the role of clinical guidelines.

Liz Smith
University of Aberdeen
Email: mes@hsru.abdn.ac.uk

I am finishing off work on my PhD about general medical practitioners' patient management decisions in depression and the role of the clinical guideline within these. The goal of the project was to discover the factors that influence their prescribing and to investigate how guideline use could be increased so as to promote clinical effectiveness. A lens model study found that GPs tend to over prescribe compared with the guidelines and place much more emphasis on thoughts of suicide and sleep disturbance than the guideline. A cluster analysis was carried out on the resulting GPs' decision policies and 3 clusters emerged which were significantly related to size of practice they worked in. GPs in the larger practices had decision policies which had much higher LME scores (when compared with guidelines) than those in medium or smaller practices. Another lens model study was carried out comparing GPs in England and Scotland and again differences were found. GPs in England prescribed at a much greater rate than those in Scotland and decision policies showed that patient treatment preference had less influence on English GPs' decisions. The results from these quantitative studies were used to generate questions for the final study which used in-depth

interviews with GPs to explore further. Many interesting data were collected which helps to explain how GPs are sometimes prevented from following guideline recommendations and fulfilling patient treatment wishes and why changing behaviour in order for it to be more compliant with guideline recommendations is so difficult. Once this PhD is completed I am starting work in the unit here, where we hope to undertake an intervention study which will use the findings of these studies.

Treatment vs. diagnosis.

Paul Sorum
Albany Medical College, Albany, NY
E-mail: sorump@mail.amc.edu

I want to alert those who read our report last year or heard my presentation in Florida that, after the meeting, Tom Stewart and Claudia Gonzalez-Vallejo redid the analyses entirely. The new version just appeared: Sorum PC, Stewart TR, Mullet E, Gonzalez-Vallejo C, Shim J, Chasseigne G, Munoz Sastre MT, Grenier B. Does choosing a treatment depend on making a diagnosis? U.S. and French physicians' decision making about acute otitis media. *Med Decis Making* 2002;22:394-402.

Further Studies of Epistemic Artifacts

Ryan D. Tweney
Bowling Green State University
E-mail: tweney@bgnnet.bgsu.edu

For the past year, most of my research effort has concentrated on further analysis of the "epistemic artifacts" used by Michael Faraday in his 1856 research on the optical properties of gold. Readers may recall that most of these are microscope slides prepared by Faraday and carefully cross-indexed to his very thorough lab notes covering the research.

Up to now, most of our efforts have centered on the difficult chemical manipulations needed to replicate these objects - we feel more like chemists sometimes than psychologists! This phase is now drawing to an end (partly as we recognize some realistic limits on what can be done) and our focus is turning to explicit cognitive analysis of the mountain of data we are facing. The "data" includes:

- (1) Faraday's own records (now digitized in several formats),
- (2) Our photographs of his specimens (about 1/5 have been photographed, but I now have the digital photography down to a routine; perhaps my next London trip will allow completion)

- (3) Records from our own replications (narrative accounts, as well as photographs and physical specimens).

How do we analyze this material? We are presently construing this as an extended protocol which manifests a variety of frame-like entities within specific "arenas" that change quite frequently (and often run in parallel). We are using techniques for representing the material based heavily on protocol analysis procedures of a "Carnegie Mellon" sort. Beyond that, we have discussed various quantitative approaches that seem promising, perhaps including Markov models of topic transitions.

Finally, we see the project as inherently Brunswikian in the sense that the task we face is how to adequately model the epistemic environment Faraday constructed and used to formulate conclusions about the topic at hand. For Brunswik, science was something like "ultimate perception." Just so.

I have been ably assisted in part or all of the work by Ryan Mears, Christiane Spitzmueller, Yanlong Sun, Chris Ayala, Neil Berg, and Robert Gibby. Mike Doherty kibbutzes on occasion, and he and I have also collaborated on a few papers (about which, see his newsletter contribution).

Persuasion, Judgment and Learning

Elise Axelrad Weaver, Ph.D.
Dept. of Social Science and Policy Studies
Worcester Polytechnic Institute
e-mail: eweaver@wpi.edu

In the last year, I moved to my new job as an assistant professor at Worcester Polytechnic Institute (www.wpi.edu). I am involved in four projects: First, I am revising an article based on my dissertation, which revisits an old debate regarding psychological reactance and impression management. The new twist is that what is monitored is an index of change in cue weights, rather than self-reported attitudes.

The second project, with Tom Stewart, is a study of the factor structure underlying judgment tasks from the Brunswikian literature, the heuristics and biases literature, and IQ and memory. We find four factors of judgment: IQ/Memory, Probability Manipulation, Learning Probabilistic Associations, and Verbal Reasoning. Three measures of accurate judgment differ from each other in their loadings across these factors, but none of these load on Verbal Reasoning. In contrast, all of the heuristics and biases tasks we measured do load on Verbal Reasoning.

The third project, with George Richardson, is a computer simulation of Hammond's

proposition that social policy thresholds will cycle under pressure from competing constituencies. This simulation uses the system dynamics technique for modeling non-linear systems with feedback.

In the fourth project, Tom Stewart and I are collecting data related to the question of whether vicarious functioning is an example of implicit learning, and whether explicit reportable knowledge aids probability learning.

Physician Diagnosis and Management of Respiratory Tract Infection

Bob Wigton
University of Nebraska Medical Center
College of Medicine
E-mail: wigton@unmc.edu

We are using data recorded from physicians and patients in clinical studies of acute respiratory tract infection to study elements of the judgement process. Do the clinical symptoms and findings differ depending on whether you ask the clinician or the patient? (Yes) Do the clinician-recorded and the patient-recorded symptoms differ in their ability to predict the correct cause of the illness? (Yes.) Are their cue weights similar when forming judgments about the possible diseases responsible for the illness? (No.) Is there wide variation in how physicians diagnose respiratory illness? (Of course). We've begun to look at the judgment modeling of the differential diagnosis and I will present some of these results at the meeting.

In a related study, we have done a judgment analysis study of the factors leading to prescription of antibiotics in a group of 95 practitioners in Denver who are taking part in a controlled trial of an intervention to reduce antibiotic prescribing. Once again, there is great variation in the strategies and the practitioners appear to be poorly calibrated with regard to antibiotic prescription (in comparison with guidelines from the CDC). We will see if the

judgment analysis results change in a repeat study of the same practitioners following the community-wide intervention this fall.

Brunswik-Symmetry

Werner W. Wittmann
University of Mannheim
wittmann@tnt.psychologie.uni-mannheim.de

We continued to capitalize in our research a lot on the symmetry principles of the lens model. We have applied these principles to different research domains with great success. A chapter titled: "Brunswik-symmetry a key concept for successful psychological research just appeared as a book chapter (unfortunately in German only). The reference is: Wittmann, W.W. (2002): Brunswik-Symmetrie, ein Schlüsselkonzept für erfolgreiche psychologische Forschung. In Myrtek, M. (Hrsg.) Die Person im biologischen und sozialen Kontext. Goettingen: Hogrefe. The chapter is full of examples from personality research, psychotherapy research and analyzing performance in complex business games additional to new methodological and conceptual issues and ideas. For those fluent in English only, can go to our homepage at: <http://www.psychologie.uni-mannheim.de/psycho2/psycho2.en.php3?page=publi/papers.en.htm&cat=publi> to find four different presentations from international conferences in PowerPoint to get a flavor of the beauty of Brunswik symmetry and how to capitalize on it in terms of prediction and explanation. How Brunswikian concepts are integrated into a larger framework concerning evaluation research and program evaluation methodology titled: "A multi-aspect approach to the evaluation of treatment quality in psychosomatic treatment" can be seen at: <http://www.psyres-stuttgart.de/kongress/program2.html>. This presentation is also coupled with my oral presentation.

Eighteenth Annual International Meeting of The Brunswik Society
Kansas City, MO, November 21st and 22nd, 2002
Westin Crown Center Hotel, Shawnee Room

Thursday, November 21st 2002

1:00 - 1:30 Late registration

1:30 - 1:45 Introductions and Welcome (Jim Holzworth, Mandeep K. Dhama & Elise Weaver)

1:45 – 3:00 Paper session 1: Applications I (Chair – Jim Holzworth)

Joshua H. Sarver, Neal V. Dawson, Susan W. Hinze, Rita K. Cydulka, & David W. Baker
- Rapid clinical decision in context: A theoretical framework to understand real-world physician decision-making.

Bob Wigton - Patient and physician models of illness in respiratory tract infection

Alex Kirlik – Vicarious functioning in action: Airline taxi navigation (and error) in the tangled web of Chicago O'Hare

3:00 - 3:15 Tea and coffee break

3:15 - 4:05 Discussion session 1

Issue - What can judgment analysis/policy capturing tell us about lay judgment? Is it appropriate to study achievement in a task that we either only perform a few times or we have no/little experience in? (Chair – Tom Stewart. Discussants – Jim Shanteau, Michael E. Doherty)

4:05 - 5:20 Paper session 2: Theoretical analysis (Chair – Nigel Harvey)

Peter Juslin, Linnea Carlsson, & Henrik Olson - Cognitive processes in multiple-cue judgment in additive and multiplicative task environments

Younho Seong - Human operators' judgment with automated decision aids: Hybrid Lens Model?

Clare Harries & Mandeep K. Dhama - Describing information search with fast and frugal models

5:20 Adjourn

6:30 Evening Group Dinner at Tomfooleries Restaurant & Bar. Sign-up on day.
(www.tomfooleries.com)

9:00 Live Kansas City blues at Tomfooleries Restaurant & Bar

Friday, November 22nd 2002

8:30 – 9:00 Tea and coffee

9:00 – 10:15 Paper session 3: Historical and methodological analysis (Chair – Tom Stewart)

Bernhard Wolf – Brunswik's concept of ecological validity

Michael E. Doherty, Ryan D. Tweney, & Yanlong Sun - Alternative hypotheses and the rule of one variable.

John R Leach, & Michael E. Doherty - effects of external representation and completely natural sampling on information selection.

10:15 – 10:30 Tea and coffee break

10:30 - 11:20 Discussion session 2

Issue - What is the role of time in judgment? Is there a place for time dynamics in Brunswikian thought? (Chair – Alex Kirlik. Discussants – Elise Weaver, Alex Wearing, James Hogge)

11:20 - 12:35 Paper session 4: Judgment Performance (Chair – Mandeep K. Dhami)

James Shanteau & Ylva Skaner - Evaluation of expert performance: comparison of two approaches

Christopher J. Anderson & Tom Stewart - Reliability and judgment competence: individual and environmental differences in multiple-cue probability learning

Elise Weaver & Tom Stewart - Is judgment more than intelligence? Four factors of judgmental skill

12:35 - 2:00 Buffet lunch and Peter Juslin – Tribute to Mats Björkman

2:00 - 3:15 Paper session 5: Applications II (Chair – Neal V. Dawson)

John Gillis & Frank Bernieri - Cross-cultural judgment of rapport

Yuko Heath, Don Hine, & Robert Gifford - Multiple regressions or hierarchical linear models?: Comparison of the two analytical approaches in decision making in a computerized fishery-management simulation

Nigel Harvey & Clare Harries - Trust in advisors: A comparison of revealed and stated measures

3:15 - 3:30 Tea and coffee break

3:30 - 4:20 Discussion session 3

Issue - How can we improve judgment? Does cognitive feedback work? (Chair – Jeryl Mumpower. Discussants - Clare Harries, Jim Holzworth)

4:20 - 4:30 Brunswik-Hammond New Investigator Prize (Awarded by Tom Stewart)

4:30 Farewell and meeting adjourned